Guided Latent Spaces for Controllable Multi-Scenario Generation in Autonomous Driving (Student Abstract)

Manasa Mariam Mammen[1], **Zafer Kayatas**[1], **Stefan Wagner**[2]
[1]Mercedes Benz AG, Sindelfingen, Germany
[2]Technical University of Munich, Heilbronn, Germany
manasa_mariam.mammen@mercedes-benz.com, zafer.kayatas@mercedes-benz.com, stefan.wagner@tum.de

Abstract

Scenario-based testing is an important approach for the development and validation of autonomous driving systems, as it enables evaluation across different driving situations. Safety-critical scenarios are especially relevant, but they occur rarely in real-world data, which creates the need for generation methods. In this paper, we present a scalable AI-based approach based on a variational autoencoder that unifies the generation of different types of critical scenarios while introducing controllability through a structured latent space. The integration of unified generation and latent space control advances AI-based scenario generation towards practical use, thereby supporting the requirements of industrial validation pipelines.

Introduction

The safety assessment of automated driving systems requires thorough testing across many possible traffic situations. Rather than driving long distances, researchers increasingly use scenario-based testing, where specific traffic situations are designed and simulated to evaluate system behavior (Webb et al. 2020). The effectiveness of this method, however, depends on the availability of high-quality scenarios that accurately reflect the variability and complexity of real-world traffic.

Recent advances in data-driven generative modeling have demonstrated the potential of AI-based approaches like variational autoencoders (VAEs) and Diffusion Models for capturing such variability directly from trajectory data (Kayatas et al. 2023; Xu et al. 2025). However, key gaps remain, since most existing methods either use separate models or work only in limited settings. This reduces their value for systematic validation. To overcome this, there is a need for a scalable and unified approach that can cover multiple safety-critical scenarios within one framework. In addition, practical use in testing requires that engineers have direct control over scenario attributes, something current methods do not provide. Taking these needs together, our work represents a first step towards an interpretable solution, where scenario generation combines unification with controllability to support large-scale testing and industrial validation.

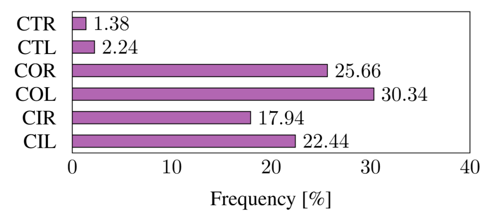

Figure 1: Probability distribution of occurrence of six considered scenarios

Method

Dataset

The dataset used is based on real-world highway measurements, combining ego-vehicle ECU signals with sensor data from surrounding traffic. It includes six safety-critical scenario types: cut-in left (CIL), cut-in right (CIR), cut-out left (COL), cut-out right (COR), cut-through left (CTL), and cut-through right (CTR). The dataset contains about 44,600 scenario trajectories, with frequencies as shown in Figure 1.

Guided Variational Autoencoder

The proposed guided variational autoencoder embeds scenario control directly into the latent space, avoiding external conditioning that can distort the underlying probability distribution. Such distortions are especially critical for failure probability estimation, where even small over- or underestimation of scenario frequencies leads to unreliable risk assessment. To solve this, one latent dimension is explicitly designated to encode the scenario class:

$$\mathbf{z} = [z_{\text{type}}, \mathbf{z}_{\text{rest}}], \tag{1}$$

where z_{type} encodes the scenario type and $\mathbf{z}_{\text{rest}}$ models the remaining trajectory variability.

The standard isotropic Gaussian prior is replaced by a class-conditional prior

$$p(\mathbf{z} \mid y) = \mathcal{N}(z_{\text{type}} \mid \mu_y, \sigma_y^2) \cdot \mathcal{N}(\mathbf{z}_{\text{rest}} \mid \mathbf{0}, \mathbf{I}), \tag{2}$$

with each class y assigned a distinct mean μ_y in the scenario dimension.

To align z_{type} with the corresponding class center, an auxiliary loss is included:

$$\mathcal{L}_{\text{align}} = \lambda_{\text{align}}\, \mathbb{E}_{\mathbf{x},y}\left[(z_{\text{type}} - \mu_y)^2\right]. \tag{3}$$

Scenario	MSE	Classification Error
CIL	5.03e-2	1.00e-2
CIR	4.17e-2	2.00e-2
COL	2.05e-2	1.00e-2
COR	6.41e-2	1.00e-2
CTL	1.20e-1	0.00e+0
CTR	9.49e-2	1.00e-2

Table 1: Validation MSE loss and classification error rates (mean values).

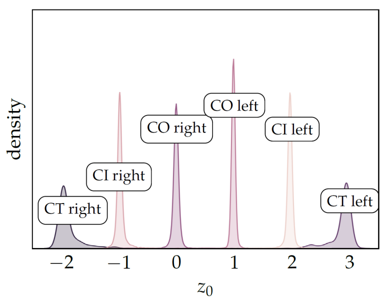

Figure 2: Learned modalities in the latent parameter z_0, which separates different scenario types.

The overall training objective then combines reconstruction, regularization, and alignment:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \beta\, \mathcal{L}_{\text{KL}} + \mathcal{L}_{\text{align}}. \tag{4}$$

This formulation preserves data fidelity while providing explicit control over scenario generation through z_{type}. The weights β and λ_{align} are selected using grid search.

Results

The model is first validated on an unseen dataset using mean squared error (MSE). Reconstruction errors remain low across all scenario types as shown in Table 1. The higher MSE error for CTL is likely due to the smaller dataset size, while scenarios with larger sample counts achieve consistently lower values.

We then analyze the latent space to verify that scenario information is properly captured. The parameter z_0 represents scenario type, and as shown in Figure 2, each scenario type forms a distinct cluster, yielding a multimodal distribution with class-specific centers that is disentangled from the other latent variables. Interestingly, longitudinal velocity, though not explicitly targeted, disentangles itself in dimension z_6, where its KDE closely matches the true velocity profiles (Figure 3). This illustrates the potential of latent parameters to capture physically interpretable attributes beyond the intended scenario dimension.

At the trajectory level, validation highlights the quality of the generated samples. Figure 4 shows that PCA projections of generated trajectories closely overlap with real data, indicating that the global structure is well preserved. In training data, the latent space forms tight scenario clusters, with cut-through trajectories lying between cut-in and cut-out, and this organization is largely maintained in generated samples, though some show slightly looser clustering and partial overlap. Scenario consistency is further confirmed by an external classifier, which yields the low error reported in Table 1. These results demonstrate realistic trajectory generation with reliable scenario control, and cluster separation could be further improved through refined latent regularization.

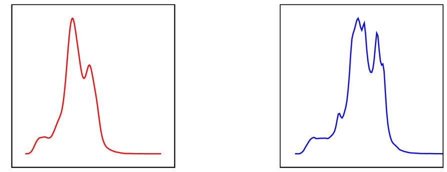

Figure 3: Latent parameter z_6 captures generalized velocity behavior. The generated trajectories (red) reproduce the real velocity profiles (blue), including peak dynamics.

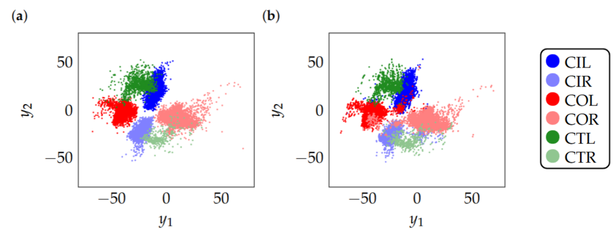

Figure 4: First two modes of PCA applied to (a) measured and (b) generated samples.

Conclusion

The work shows that complex driving scenarios can be generated realistically while remaining controllable through interpretable latent parameters of AI models. This controllability makes the approach both practical and effective for testing and validation of autonomous driving systems. Future work will focus on improving explainability and grounding the latent space in more physical parameters to further enhance interpretability and reliability.

References

Kayatas, Z.; Bestle, D.; Bestle, P.; and Reick, R. 2023. Generation of Realistic Cut-In Maneuvers to Support Safety Assessment of Advanced Driver Assistance Systems. *Applied Mechanics*, 4(4): 1066–1077.

Webb, N.; Smith, D.; Ludwick, C.; Victor, T.; Hommes, Q.; Favaro, F.; Ivanov, G.; and Daniel, T. 2020. Waymo's safety methodologies and safety readiness determinations. *arXiv preprint arXiv:2011.00054*.

Xu, C.; Petiushko, A.; Zhao, D.; and Li, B. 2025. Diffscene: Diffusion-based Safety-Critical Scenario Generation for Autonomous Vehicles. *In Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 39, No. 8, pp. 8797–8805.