Generating Traffic Scenarios via In-Context Learning to Learn Better Motion Planner

Alexander (Aizierjiang Aiersilan)

University of Macau

View Project Materials

Abstract

Motion planning is a crucial component in autonomous driving. State-of-the-art motion planners are trained on meticulously curated datasets, which are not only expensive to annotate but also insufficient in capturing rarely seen critical scenarios. Failing to account for such scenarios poses a significant risk to motion planners and may lead to incidents during testing. An intuitive solution is to manually compose such scenarios by programming and executing a simulator (e.g., CARLA). However, this approach incurs substantial human costs. Motivated by this, we propose an inexpensive method for generating diverse critical traffic scenarios to train more robust motion planners. First, we represent traffic scenarios as scripts, which are then used by the simulator to generate traffic scenarios. Next, we develop a method that accepts user-specified text descriptions, which a Large Language Model (LLM) translates into scripts using **in-context learning**. The output scripts are sent to the simulator that produces the corresponding traffic scenarios. As our method can generate **abundant safety-critical traffic scenarios**, we use them as synthetic training data for motion planners. To demonstrate the value of generated scenarios, we train existing motion planners on our synthetic data, real-world datasets, and a combination of both. Our experiments show that motion planners trained with our data significantly **outperform those trained solely on real-world data, showing the usefulness of our synthetic data and the effectiveness of our data generation method**.

Motivations

- High cost of collecting real-world data and limitations of current datasets.
- Risk of incidents due to unaccounted safety-critical scenarios.
- Substantial human costs of manually composing scenarios.

Proposal

- An inexpensive method for generating diverse critical traffic scenarios.
- Representing traffic scenarios as scripts for simulators.
- Using LLMs to translate user-specified text descriptions into scripts.
- Generating abundant safety-critical traffic scenarios for synthetic training data.
- Collecting data from the physics-based simulator to augment/replace real-world datasets.

Contributions

Scenario generation has traditionally been manual and labor-intensive. However, advancements in LLMs allow for efficient AI-driven generation of specific traffic scenarios. This study builds on prior research and makes the following key contributions:

1. A universal, general, and cost-effective framework, "AutoSceneGen", is proposed to automatically enhance the heterogeneity of traffic scenarios through scenario descriptions, thereby accelerating the simulation and testing process.
2. AutoSceneGen leverages in-context learning (ICL) of LLMs, eliminating the need for training or fine-tuning generative models for scenario generation tasks.
3. The scenarios generated by AutoSceneGen were demonstrated to produce better datasets, leading to improved training results for motion planners.
4. AutoSceneGen automatically categorizes scenarios by their descriptions, removing the need for downstream annotation and aiding motion planner training in open-world environments.
5. AutoSceneGen is modular with dynamic components, enabling easy replacement of its generative model and simulation engine for scenario generation and data collection.

AutoSceneGen Framework

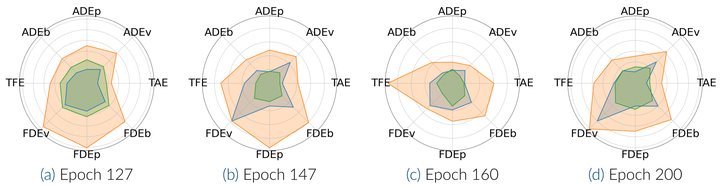
AutoSceneGen consists of key components for processing scenario descriptions, which can be provided by the user or extracted from images using a vision-language model. A filtering process ensures simulator compatibility by replacing incompatible terms with appropriate alternatives.

Figure 1. **Architecture Overview**. It begins with the user inputting a scenario description, which is managed by the *Exception Handler* to block adversarial or irrelevant inputs, ensuring the framework operates within scope and prevents downstream issues. The *Filter* processes the description, replacing simulator-incompatible terms with those aligned to the simulator's documented APIs. The filtered description (*Desc.'*) is combined with pre-constructed ICL exemplars, which can be zero-shot, one-shot, or few-shot in category, depending on the LLM's familiarity with the simulator's APIs and the complexity of the scenario. The LLM generates a response containing scenario configurations, often accompanied by explanations and comments. The *Validator* verifies each API call for compatibility, replacing unsupported terms with suitable alternatives (e.g., replacing "storm," unsupported in CARLA, with "rain") or ignoring them to prevent errors. This ensures all calls align with the simulator's capabilities, enabling execution of the final configuration file. The simulator runs the scenario, with the final step depicting the interaction between the real world and the virtual environment, while data collection can take place either inside the simulator or externally.

Results

This study addresses the challenge by leveraging LLMs' ICL capabilities to generate tailored configurations for rare scenarios, streamlining the ideation and scenario creation processes. Figure 2 shows a rare scenario generated with this approach.

Figure 2. Images captured at four distinct timestamps and locations, corresponding to input scenario description: "In downtown area, during a drizzly noon, there are vehicles malfunctioning windshield wipers and some of the vehicles' doors are open. Some vehicles exhibit negligent driving behavior, compromising visibility in wet conditions. There are 10 pedestrians on the road, with 50% of the pedestrian running. No one was hurt and no accident happened since all the vehicles except the malfunctioning one obeyed the traffic rules."



(a) Epoch 127 (b) Epoch 147 (c) Epoch 160 (d) Epoch 200

Figure 3. The comparison of all metrics between the datasets collected via AutoSceneGen (Blue), ApolloScapes (Orange; A.S.), and the combination of the two datasets (Green) across different epochs is shown. While the dataset collected purely from AutoSceneGen outperforms A.S. in some epochs, such as epoch 127, the combination of AutoSceneGen and A.S. demonstrates better overall results. Due to the distinct distribution of traffic participants in the two datasets, Figures (b), (c), and (d) show sharper peaks for FDE-vehicle and ADE-vehicle. However, the combination of the two datasets achieves reasonable values overall. In this experiment, A.S. has a total of 3,917 frames, AutoSceneGen has 17,919 frames, and the combined AutoSceneGen + A.S. has 27,605 frames.

Comparisons

Without modifying the original trajectory prediction network, our dataset achieved superior results with reduced displacement error for each traffic participant type, as shown in Table 1. In various epochs, the dataset collected from AutoSceneGen demonstrated the highest accuracy in trajectory prediction, as illustrated in Figure 3-(a). Moreover, combining our dataset with ApolloScapes improved overall performance, enhancing all trajectory prediction metrics by incorporating diverse scenarios and extensive data, as depicted in Figures 3(b), (c), and (d).

Dataset	Method	TAE	ADEv	ADEp	ADEb	TFE	FDEv	FDEp	FDEb
A.S.	TrafficPredict	0.085	0.080	0.091	0.083	0.141	0.131	0.150	0.139
A.S. + Ours	TrafficPredict	0.053	0.085	0.058	0.065	0.076	0.114	0.092	0.094
Ours	TrafficPredict	0.033	0.088	0.020	0.047	0.058	0.135	0.037	0.077
TRAF	TraPHic	5.63	N/A	N/A	N/A	9.91	N/A	N/A	N/A
A.S.(Reproduced)	TraPHic	5.10	3.62	1.02	4.49	2.81	6.73	1.88	8.44
A.S. + Ours	TraPHic	1.30	1.69	0.42	0.90	2.10	2.82	0.67	1.39
Ours	TraPHic	0.27	0.14	0.19	0.44	0.40	0.21	0.30	0.62

Table 1. The comparison results of the ApolloScapes dataset, collected from AutoSceneGen (Ours), and the two datasets combined are presented. The term "ApolloScapes Dataset" is abbreviated as "A.S." in the table. Lower metrics indicate better performance. We generated 17,919 examples; the official training set of A.S. contains 94 examples. We used the original method proposed in TrafficPredict to train the planner. For TRAF and another evaluation, we used TraPHic. While the results are not as good as the results under TrafficPredict, under the method TraPHic there are still huge improvements thanks to the substitution of TRAF and ApolloScapes to the dataset collected via AutoSceneGen. "ADE" stands for Average Displacement Error, and "FDE" stands for Final Displacement Error, with suffixes "v," "b," and "p" representing vehicle, bicycle, and pedestrian, respectively.

Dataset	Method	ADE	FDE
NGSIM	Pihgu	0.88	1.96
Ours	Pihgu	7.98	15.43
NGSIM train-set + Ours	Pihgu	0.84	1.87
ETH/UCY	Pihgu	1.10	2.24
Ours	Pihgu	1.48	2.70
ETH/UCY train-set + Ours	Pihgu	0.79	1.50
VIRAT/ActEV	Pihgu	14.11	27.96
Ours	Pihgu	16.05	31.09
VIRAT/ActEV + Ours	Pihgu	15.32	29.65

Table 2. The comparison results of the NGSIM dataset, the dataset collected from AutoSceneGen (ours), and the combination of the two datasets are shown. When replacing the NGSIM dataset with ours, the ADE and FDE values are much higher (worse) than when using the original NGSIM dataset. However, when we combine the two datasets—NGSIM and ours—the ADE and FDE decrease and outperform the results obtained from using the NGSIM dataset alone, indicating that the original NGSIM dataset is augmented by our dataset collected via AutoSceneGen. A similar observation was made with the ETH/UCY dataset. [Insights are detailed in the paper.]

References

[1] Alinezhad et al. Pishgu: Universal path prediction network architecture for real-time edge cps. In *ICCPS 2023*, page 88, 2023.
[2] Chandra et al. Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions. In *CVPR*, page 8483, 2019.
[3] Dosovitskiy et al. Carla: An open urban driving simulator. In *CoRL*, pages 1–16. PMLR, 2017.
[4] Huang et al. The apolloscape dataset for autonomous driving. In *CVPR workshops*, pages 954–960, 2018.
[5] Lerner et al. Crowds by example. In *Computer graphics forum*, volume 26, pages 655–664. Wiley Online Library, 2007.
[6] Liu et al. Large language models are few-shot health learners. *arXiv preprint arXiv:2305.15525*, 2023.
[7] Ma et al. Trafficpredict: Trajectory prediction for heterogeneous traffic-agents. In *AAAI*, volume 33, page 6120, 2019.
[8] Pellegrini et al. You'll never walk alone: Modeling social behavior for multi-target tracking. In *ICCV 2009*, pages 261–268. IEEE, 2009.
[9] Sangmin et al. A large-scale benchmark dataset for event recognition in surveillance video. In *CVPR 2011*, page 3153. IEEE, 2011.