GenSUMO: GenAI-Creation of Critical Scenarios for Autonomous Vehicle Testing

Sergio Augusto Angelini
Department of Control and Computer Engineering
Politecnico di Torino
Turin, Italy
sergioaugustoangelini@gmail.com

Diego Gasco
Department of Control and Computer Engineering
Politecnico di Torino
Turin, Italy
diego.gasco@polito.it

Claudio Casetti
Department of Control and Computer Engineering
Politecnico di Torino
Turin, Italy
claudio.casetti@polito.it

***Abstract*—Recent advancements in autonomous vehicle research highlight the importance of Machine Learning (ML) models in tasks like motion planning, trajectory prediction, and emergency management. To support AI development, we propose a novel approach for generating on-demand datasets using the Simulator of Urban Mobility (SUMO) and a Generative Adversarial Network (GAN). Our method focuses on capturing critical events such as sudden pedestrian crossings, near-misses, and collisions, providing essential data to improve vehicle models' responses to emergency situations.**

***Index Terms*—Autonomous vehicles, Critical environments, Vehicular simulation, SUMO, AI, GAN**

I. Introduction

Artificial Intelligence significantly enhances autonomous vehicles by leveraging advanced Machine Learning (ML) models. However, these models face several challenges, particularly the need for large and diverse datasets to effectively manage critical situations. Unfortunately, creating specific types of emergencies using traffic simulators like SUMO and CARLA can be quite difficult. To address this issue, we present GenSUMO [1], a framework designed to generate critical scenarios in SUMO, facilitating the creation of targeted datasets for ML algorithms.

II. Related Works

In the automotive industry, the frameworks based on Machine learning are pivotal for motion planning and collision avoidance systems. These systems depend on large datasets [2] [3] encompassing diverse scenarios; however, the costs and challenges associated with manual data collection continue to pose significant obstacles. In [4] and [5], authors highlight the possibility of generating synthetic datasets using Generative Artificial Intelligence (GenAI) solutions. Starting from these achievements, we employ GenAI to create synthetic datasets specifically for critical situations. Generative Adversarial Neural Networks (GAN) are used to develop emergencies such as vehicle-to-vehicle and vehicle-to-pedestrian collisions. In particular, GANs use a Generator and Critic architecture, with the Generator creating synthetic data and the Critic distinguishing between real and synthetic data, aiming for the synthetic data to be indistinguishable from the real.

III. Categorize Critical Situations for Dataset Building

Vehicles and pedestrians often face critical situations caused by traffic violations or distractions. To describe collisions accurately, we develop a gravity measure based on the type of collision and the resulting damage. This allows us to build a dataset of labeled critical events categorized by gravity. We use the Abbreviated Injury Scale (AIS) [6] to link injuries with life-threatening risks. However, as AIS is too detailed for classifying patient injuries, we instead use the Maximum Abbreviated Injury Scale (MAIS), which highlights the most severe injury and is commonly applied in probabilistic models for road safety assessment. The work in [7] suggests that the speed difference before and after a collision (ΔV) is a strong predictor of crash severity, but it is not enough as detailed in [8]. This study shows a system for vehicle-to-vehicle crashes assigns a MAIS3+ rating for serious harm and it incorporates both ΔV and vehicle impact sections based on angle (Near side, Far side, Front, Rear), using four lognormal regressions to model injury probability distributions. To account for minor variations in ΔV values, we introduce two thresholds (Θ_1 and Θ_2) to the lognormal distributions in [8] for each collision type, creating three severity levels: minor, serious, and critical collisions. Additionally, we incorporate vehicle-to-pedestrian collisions without differentiating impact positions, resulting in 15 categories: 12 from four vehicle-to-vehicle collision types (Near side, Far side, Front, Rear) and 3 from vehicle-to-pedestrian impacts. Using SUMO to track vehicle states, we can gather pre-collision or braking data and categorize events accordingly, applying this method across any SUMO scenario.

IV. GAN for Scenario Generation

Our method employs a Conditional Wasserstein GAN (CGAN), where the generator is constrained to produce data for specified classes within the classification system. The framework outputs a matrix representing agents' (x, y) positions and velocity v, depicting snapshots up to 10 seconds before critical events. Generated data must follow road network constraints, such as lane positioning and speed limits. To enforce this aspect, we incorporate a Physics-Informed

Neural Network and a custom loss to penalize violations. The final loss function consists of the critic's score for real data conditioned on labels, the critic's score for synthetic data, a gradient penalty for stable training, and the regularization term to ensure compliance with physical and road network constraints. Results show that combining these losses and regularization terms is essential for convergence and generating valid synthetic data.

V. Experiments and Results

The framework aims to generate configuration files for SUMO that lead to specific collision events based on the emergency classification method described earlier, using a road network with four lanes converging at a regulated intersection where pedestrians can move freely and cross at designated crosswalks. SUMO enables the assignment of probabilities for vehicles to disregard right-of-way rules, potentially causing collisions whose severity can be evaluated using our classification system. We set a minimum threshold of 50 vehicles and 50 pedestrians for saving snapshots, focusing on the closest agents to the intersection to capture relevant dynamics. To address class imbalance in emergencies, we sample data with a probability inversely proportional to each class's frequency, ensuring uniform distribution in training batches. An Adam optimizer is adopted with a learning rate of 0.00005 and a gradient penalty coefficient of 10, reducing the learning rate by 10% each epoch while updating the critic 20 times per generator iteration. We statistically evaluate our framework by estimating the empirical likelihood of these events using a 95% Wilson score confidence interval, compared to a control group. The control group comprises 2,740 instances generated with fixed vehicle and pedestrian counts of 50, as used in our dataset. The evaluation covers both macro categories (vehicle-vehicle and vehicle-pedestrian collisions) and specific event classes. Figure 1 shows our framework generates 0.12 more vehicle-vehicle and 1.06 more vehicle-pedestrian collisions per simulation than the control group. The topology of our SUMO traffic network allows the generation of just a subset of these collision classes: Vehicle-Vehicle Lateral Driver Low (V-V-LDL), when the collision occurs on the driver side; Vehicle-Vehicle Lateral Low (V-V-LL), when the collision occurs on the passenger side; Vehicle-Pedestrian Low (V-P-L); Vehicle-Pedestrian Medium (V-P-M). The system outperforms in generating V-V-LDL, V-V-LL, and V-P-M collisions but struggles with the V-P-L category, likely due to class imbalance. Expanding the dataset should effectively address these discrepancies.

VI. Conclusions and Acknowledgments

We developed a methodology to classify critical situations in vehicle-to-vehicle and vehicle-to-pedestrian collisions, enabling the analysis of data from both simulated and real-world scenarios. This labeling process supports the creation of datasets for Machine Learning models designed to predict autonomous vehicle behaviors. To address the growing need for training data, we proposed a framework that generates

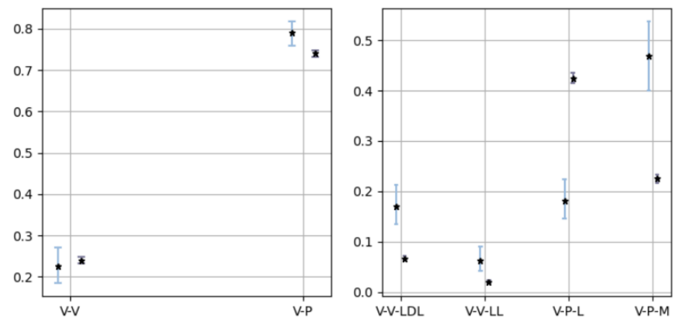

Fig. 1: Empirical probabilities (black stars) and corresponding Wilson confidence intervals. Each graph contains results for both GAN-generated data (left side) and control data (right side). The graph on the left shows general vehicle-vehicle and vehicle-pedestrian collision events, while the graph on the right presents results for the class-specific testing procedure.

synthetic simulations using Conditional Generative Adversarial Networks combined with Physics-Informed Systems. Our initial implementation uses SUMO traces labeled by criticality, but the framework is flexible and can adapt to other simulated or real data sources. These contributions offer essential tools for training and validating Machine Learning models, ensuring accurate and reliable responses from autonomous vehicles.

This work was funded in part by the project PNRR-NGEU which has received funding from MUR - DM 117/2023 and in part by the European Union through the project CONNECT under grant agreement no. 101069688. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] D. Gasco, "Gensumo," https://github.com/Diegomangasco/GenSUMO, 2024.

[2] U.S. Department of Transportation Federal Highway Administration, "Next generation simulation (ngsim) vehicle trajectories and supporting data," Dataset, 2016, provided by ITS DataHub through Data.transportation.gov. Accessed YYYY-MM-DD. [Online]. Available: http://doi.org/10.21949/1504477

[3] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 2118–2125.

[4] D. Dauner, M. Hallgarten, T. Li, X. Weng, Z. Huang, Z. Yang, H. Li, I. Gilitschenski, B. Ivanovic, M. Pavone, A. Geiger, and K. Chitta, "Navsim: Data-driven non-reactive autonomous vehicle simulation and benchmarking," *arXiv*, vol. 2406.15349, 2024.

[5] K. Chitta, D. Dauner, and A. Geiger, "Sledge: Synthesizing simulation environments for driving agents with generative models," vol. 2403.17933, 2024.

[6] J. D. States, "The abbreviated and the comprehensive research injury scales." *STAPP Car Crash Journal. 13*, 1969.

[7] M. Mackay, "Injury and collision severity," *12th Stapp Car Conf.*, pp. 207–219, 02 1968.

[8] C. Jurewicz, A. Sobhani, J. Woolley, J. Dutschke, and B. Corben, "Exploration of vehicle impact speed – injury severity relationships for application in safer road design," *Transportation Research Procedia*, vol. 14, pp. 4247–4256, 12 2016.