Mock On-Road Test for Autonomous Vehicles in Test Tracks

Shuo Feng
Assistant Research Scientist
University of Michigan Transportation Research Institute
University of Michigan, Ann Arbor

April 13th, 2022
CCAT Global Symposium
Transportation Revolutions in the Last 100 Years

**1908**
- Ford Model T
  - Mass production on moving assembly lines
  - The first affordable automobile

**1956**
- National Highway System
  - Railway system became obsolete
  - Travelers either on the wheels or on the flight

**Now & Future**
- Smart Transportation System
  - Connected
  - Automated
  - Electrified
Biggest technical challenge for AVs

Limitations of Deep Learning

- The Long Tail Problem - Rare events require an exponential amount of data to train and validate.

- Rare events include *life or death* situations.

- "When you are 90% done, you still have 90% to go."
  
  -- Sacha Arnoud, Director of Engineering, Waymo

Images from Waymo
Purposes of AV testing

Behavior Competency and Safety Performance Evaluation

$10^{-6}$ Accidents Per Mile

Long Tail of Events

Normal events

Rare events

Scenario Identification & Training
How to test an AI-based driving system?

- There are no consensus nor standard procedures on how to test and evaluate AVs.

- The AI-based agent, which is usually a black box to external users, limits the use of traditional logic-based software verification and validation techniques.

- The prevailing state-of-the-art approach for AV testing uses the agent-environment framework, through a combination of software simulation, closed-track testing, and on-road testing.
Existing testing methods

Simulation
Low Fidelity

Test Track
Lack of Traffic

On-Road
Time & Cost Expensive
How many miles to validate an AV?

1.1 deaths per 100 million miles (According to NHSTA in 2019)

To prove AV’s are 20% better than human drivers using a fleet of 100 AV’s driving 25 mph:

Avoiding Crashes – 28M miles (1.3 years)
Avoiding Injuries – 170M miles (7.6 years)
Avoiding fatalities – 5B miles (225 years)

Testing a human driver vs. testing an AV in tracks

- Road Test
- Track Test
- Written Test
- Physical Exam

Mock On-Road Test:
- Highway Driving
- Urban Driving

Behavioral Competency Test in Dynamic Environment:
- Lane Change
- Car Following

Behavioral Competency Test in Static Environment:
- Maintaining Speed
- Obstacle Avoidance

Perception and Control Test:
- Sensor Test
- Localization Test

Motor Vehicle Test:
- Collision Test
- Emission Test
- Durability Test
Three challenges of mock on-road testing

• Challenge #1: Lack of background traffic

  • A test track is usually empty and lack of background traffic.
  • There may exist a few proxy objects, which suffer from the limited numbers and controllability.

VS.

• However, the key to the safety performance of AVs is the interaction with the background traffic.
Three challenges of mock on-road testing

• Challenge #2: How to determine the behaviors of background traffic?

Naturalistic driving environment (NDE) modeling:
• “Curse of dimensionality” caused by the complexity of the driving environment

Weather

Road Infrastructure

Road Users

Maneuvers
Three challenges of mock on-road testing

• Challenge #2: How to determine the behaviors of background traffic?

Naturalistic driving environment (NDE) modeling:
• “Curse of dimensionality” caused by the complexity of the driving environment
• The NDE should follow the same distribution with the realistic traffic distribution
Three challenges of mock on-road testing

• Challenge #3: How to accelerate the testing process?

Due to the rareness of safety-critical events, hundreds of millions of miles would be required in NDE to demonstrate the safety performance of AVs.

The automated vehicle must give way to the emergency vehicle even it has the right of way. A low probability but potentially safety-critical event.

Summary of the challenges

Lack of background traffic

The lack of background traffic severely limits the testing scenarios in test tracks.

High-dimensional spatiotemporal stochastic environment

Variables that define the environment are high dimensional, which can cause the “curse of dimensionality”. In addition, the underlying distribution needs to be consistent with that of the real world NDE.

Rareness of events

The rareness of events can lead to the intolerable inefficiency issue for testing.
SAFE-TEST: Safe AI Framework for Trustworthy Edge Scenario Tests

- **Augmented Reality (AR):** Using simulated vehicles as background traffic
- **Naturalistic and Adversarial Driving Environment (NADE):** Improving the fidelity of simulation through naturalistic driving environment (NDE) modeling and accelerating the testing through adversarial scenario generation
Values of SAFE-TEST

1 Test Track Mile = 5000+Road Miles

Cost 50 – 100x Less

• Integrate best-in-class solutions from tool chain to public road through smart mobility test center and partner ecosystem

1. Reduce development costs
2. Compress development cycle
Solution #1: Augmented reality testing platform

Infrastructure-based AR testing platform

Cloud Computing Environment
- SUMO with NADE
- AR data interface

ACM Test Facility
- RSU

Automated Driving System
- DSRC
- OBU

Vehicle-based AR testing platform

ACM Test Facility

Automated Driving System

Onboard Computing Environment
- SUMO with NADE
- AR data interface

(Feng et al., Accident Analysis & Prevention, 2020)
Cut-in Case Study: Field Test in Mcity

(Feng et al., Accident Analysis & Prevention, 2020)
Augmented perception

• Augment the real-world videos with the virtual vehicles generated by NADE.

Real-world videos from Ford AV dataset

Augmented videos by our solution
Solution #2: Framework of NADE modeling

Test Objectives
E.g., highway driving performance

Naturalistic Driving Datasets

Data Processing
Naturalistic Driving Environment (NDE) Modeling
To reproduce the real-world traffic environment

Naturalistic Adversarial Driving Environment (NADE) Modeling
To accelerate the AV validation by sparse adversarial adjustments to NDE.

Trustworthy Scenarios
Most of existing models are developed for traffic flow simulation, not for safety evaluation.
A data-driven NDE modeling framework

• The goal of NDE is to reproduce the real-world traffic environment.
The key idea of NADE is to train the background vehicles in the NDE to learn *when* to execute *what* adversarial maneuver, while ensuring *unbiasedness* and improving *efficiency*.

The theory behind is that the use of the importance sampling method with the small subset of critical variables, while applying the crude Monte Carlo method with the remaining variables can satisfy both the unbiasedness and efficiency requirements.

(Feng et al., Nature Communications, 2021)
Rare event estimation with importance sampling

- **NDE Sampling:**

\[
P(A) \approx \frac{1}{n} \sum_{j=1}^{n} P(A|X_j) \approx \frac{m}{n}, \; X_j \sim P(X)
\]

- **Importance sampling:** construction of a new sampling distribution

\[
P(A) = \sum_{X \in X} P(A|X)P(X) = \sum_{x \in X} \frac{P(A|X)P(X)}{q(X)}q(X) \approx \frac{1}{n} \sum_{j=1}^{n} \frac{P(X_j)}{q(X_j)} P(A|X_j), \; X_j \sim q(X)
\]

Construct an importance funtion \( q(X) \) so that:

1. \( P(A) \) is an unbiased estimation
2. The number of required tests (or testing miles) is smaller (fewer) than public road tests

The construction of importance function leads to the development of **naturalistic and adversarial driving environment (NADE)**

(Feng et al., Nature Communications, 2021)
Procedure of NADE Generation

For any time step:

1. Criticality calculation
2. Exposure Frequency
3. Maneuver Challenge
4. Criticality of BVs
5. Identification of POV
6. Sampling maneuvers from the importance function
7. One time step simulation

(Feng et al., Nature Communications, 2021)
Accelerating AV Testing With NADE

Testing results in NADE can be converted into equivalent testing results in NDE, but with much fewer testing miles, because the frequency of adversarial challenges to ADS in NADE is much higher than that in NDE.

\[
\frac{\text{# of long tail events per mile in NADE}}{\text{# of long tail events per mile in NDE}} = X
\]

\[
\text{1 mile in NADE} \approx X \text{ miles in NDE}
\]

(Feng et al., Nature Communications, 2021)
Case study of highway driving environment

Safety Pilot Model Deployment (SPMD) Database

- About 140 vehicles equipped with Mobileye and DAS

To obtain the same accurate estimation, NADE requires about $2.32 \times 10^4$ miles, while NDE requires about $1.41 \times 10^8$ miles – an acceleration of 6,000 in this case study.
Sparse adversarial adjustments

- NADE generates very similar distributions as NDE (naturalistic), but much more dangerous scenarios with small distances and TTC (adversarial).

- We investigate the adjustment frequency of BVs’ maneuvers in NADE: we only adjust about 1.5% maneuvers of the environment, which is very sparse and thus keep the environment naturalistic.

(Feng et al., Nature Communications, 2021)
More valuable events

- We compare the **events** encountered by the AVs in NDE and NADE:

- BV cut in
- BV hard brake
- Lane conflict
- AV lane change

![Diagram](image)

(Feng et al., Nature Communications, 2021)
Unbiased accident types

- We further investigate the unbiasedness of accident types:

(Feng et al., Nature Communications, 2021)
Long tail events in NADE

Example 1
Systematic safety performance analysis

Counterfactual Simulation

Long-tail Event Augmentation

Safety Metrics

Accident Type Analysis

Reason’s Swiss cheese model

(Source: NHTSA, 2019)

Accident Responsibility Analysis

(Source: Sheridan, 2008)

Control Errors

Perception Errors

Localization Errors

Strengths and Weakness of AVs

(Source: Sheridan, 2008)
Implementation at ACM
Implementation at ACM
Implementation at ACM

Hardware stack: UM OpenCAV Platform

- By-wire Dataspeed
- Velodyne 32 channel LiDAR 360°
  - Horizontal, +10°/−30° Vertical
  - 100-120 meters, Dual Returns
- DSRC COHDA MK5
  - (3-27 Mbps)
- 3 IBEO LiDAR
  - 2 front, 1 rear
  - 200 meters
  - 110° Horiz., 3.2° Vertical
- Mobileye 560
- PointGrey
- 4 Delphi SRR
  - 4 corners, 0.5-80m/150°
- 1 Delphi ESR
  - Front, 60m/90° 175m/20°
- XTIK 900 MHz radio modems
- OTX GPS RTK GPS RT3003
- Xsens MTi GPS/IMU

Software stack: Autoware.AI

Robot Operating System (ROS)
Demonstration at ACM: Near crash BV cut-in AV
Demonstration at ACM: Crash BV cut-in AV

Crash: BV cut-in AV
Demonstration at ACM: Near-crash AV Lane Change
Demonstration at ACM: Highway Merging
Demonstration at ACM: Deer Accident
Demonstration at ACM: Overtake Truck

Overtake Truck
Demonstration at ACM: Augmented Rain
SAFE-TEST use procedure

**PREPARATION**

Step 1: Determine testing objectives
- Scenario category or ODD
- Test platform:
  - Simulation
  - Test track: with infrastructure- or vehicle-based AR

Step 2: Prepare the AV under test
- For simulation: AV model
- For test track: AV needs to be able to receive BSM and output high-accuracy localization information.

**SCENARIO GENERATION**

Step 3: Select the NDD data
- Opt-1: UM datasets
- Opt-2: customer’s datasets. Ego vehicle and surrounding vehicles: position, lateral distances, speed, and acceleration, etc.

Step 4: NADE Generation
- 4.1: NDE data processing
- 4.2: NADE generation by training intelligent background vehicles.

**AV TESTING**

Step 5: Perform the test
- 5.1: Simulation test or field test in ACM.
- 5.2: Stop the test when converges.

Step 6: Results analysis
- 6.1: Safety metric calculation.
- 6.2: Statistical performance analysis.
Representative Publications


Acknowledgements

• Collaborators

Dr. Henry Liu
Xintao Yan
Haowei Sun
Haojie Zhu
Shengyin Shen

• Research sponsors
Contact Information

Shuo Feng, Ph.D.
University of Michigan
Transportation Research Institute
2901 Baxter Road
Ann Arbor, MI 48109-2150 USA
Email: fshuo@umich.edu