SALUS: A Novel Data-Driven Monitor that Enables Real-Time Safety in Autonomous Driving Systems

Bohan Zhang
bzhang22@uiowa.edu

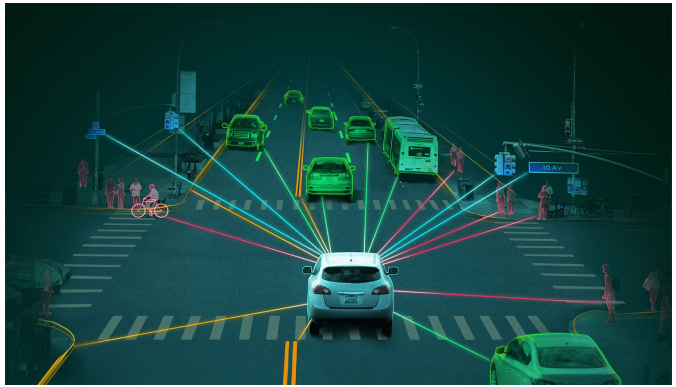
Yafan Huang
yafan-huang@uiowa.edu

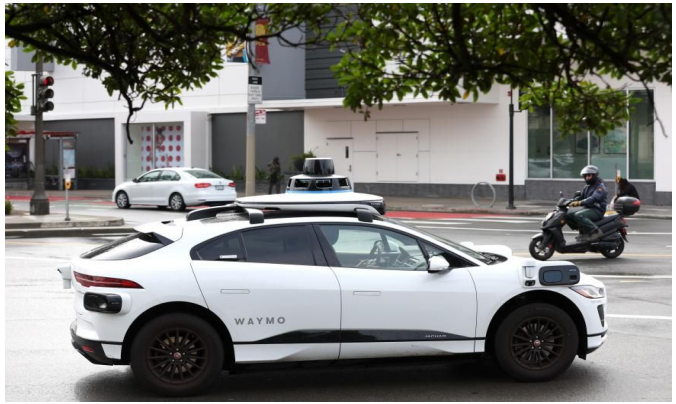
Guanpeng Li
guanpeng-li@uiowa.edu

IOWA

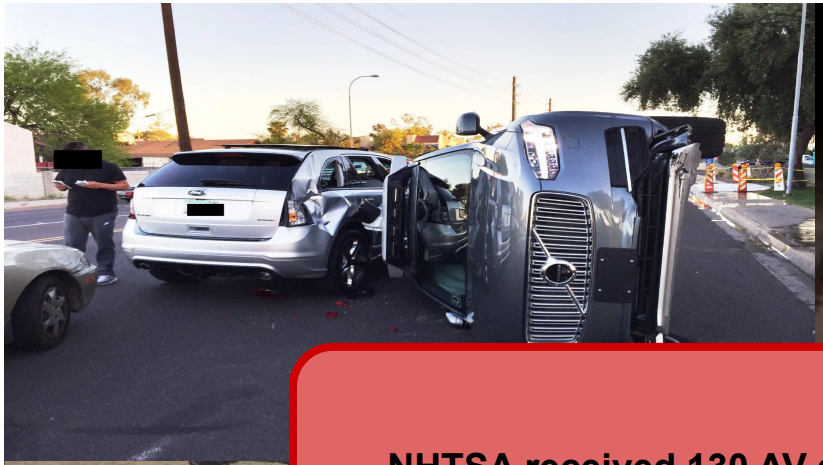
Autonomous Vehicle (AV)

- Every 10 mins, about 26 people will die in automobile accidents globally
  - >94% is drivers' fault that cause the accidents
- AV holds significant potential to improve productivity and quality of life
  - Level-4 autonomous driving system (ADS) requires no human intervention
- By 2019, over 1,400 AVs are in testing by 80+ companies across the US on roads
  - Operate with human-driven cars on public roads
  - Market estimate of AV will reach $325.9 billion by 2030
  - Number of AV will reach 33 million by 2040

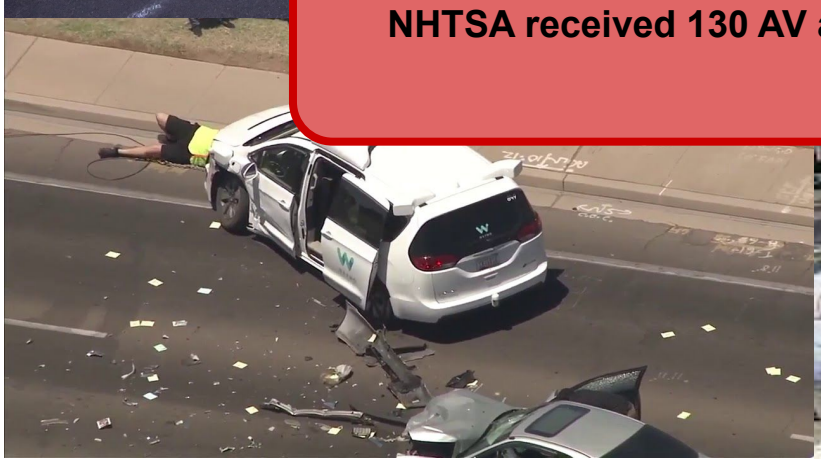




AV Accidents



NHTSA received 130 AV accidents from 2021-2022



Existing Methods

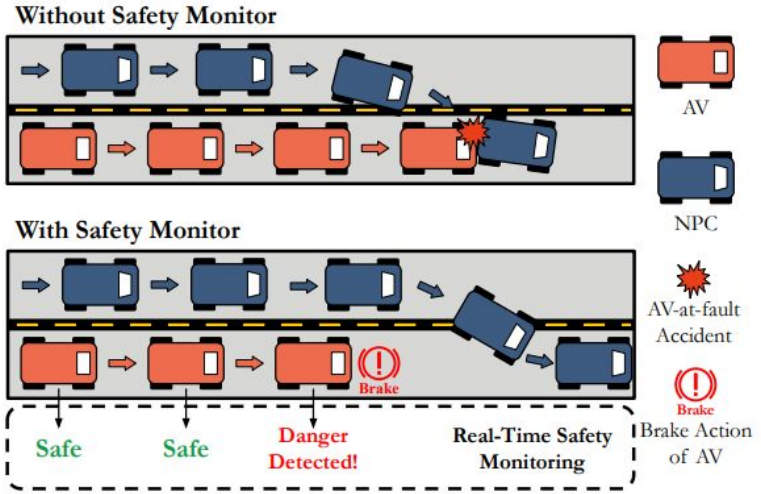
- Road Testing
  - Require million of millages and energy
  - Not efficiency (requrie human to monitoring)

- Simulation-Based Software Fuzzing
  - Hard to localize the error (huge code base,

    more than 400000 line of codes)
  - Fix one bug may cause more

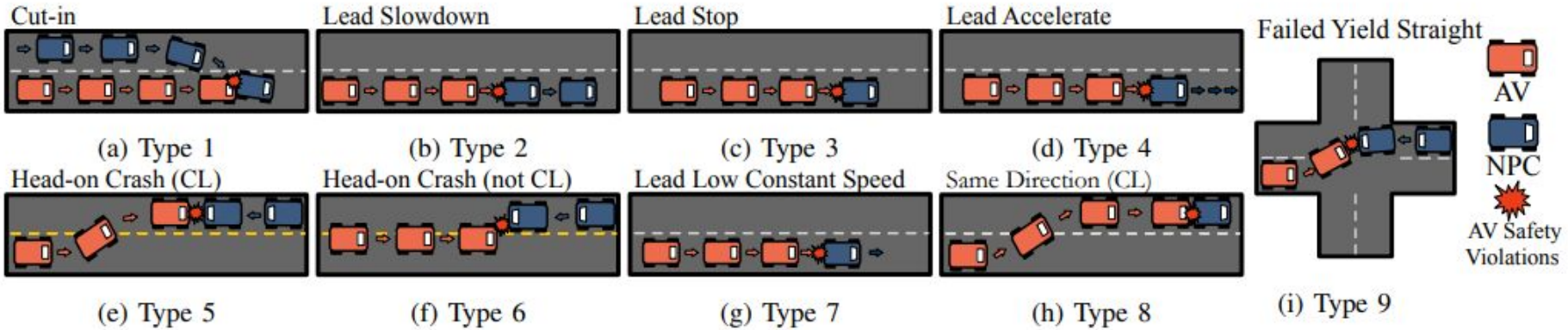
Our Solution

- Key insights
  - NPC (surrounding vehicle)’s trajectoris which will lead the AV safety violations fall into identifiable patterns
- ML-based real-time safety monitor
  - Without debugging the code of ADS
  - Real-time dectection for safey violation
  - Mitigation when danger dectected



Enable Maxium safety

Initial Study



NHTSA
www.nhtsa.gov

Baidu 百度 | apollo

SVL SIMULATOR

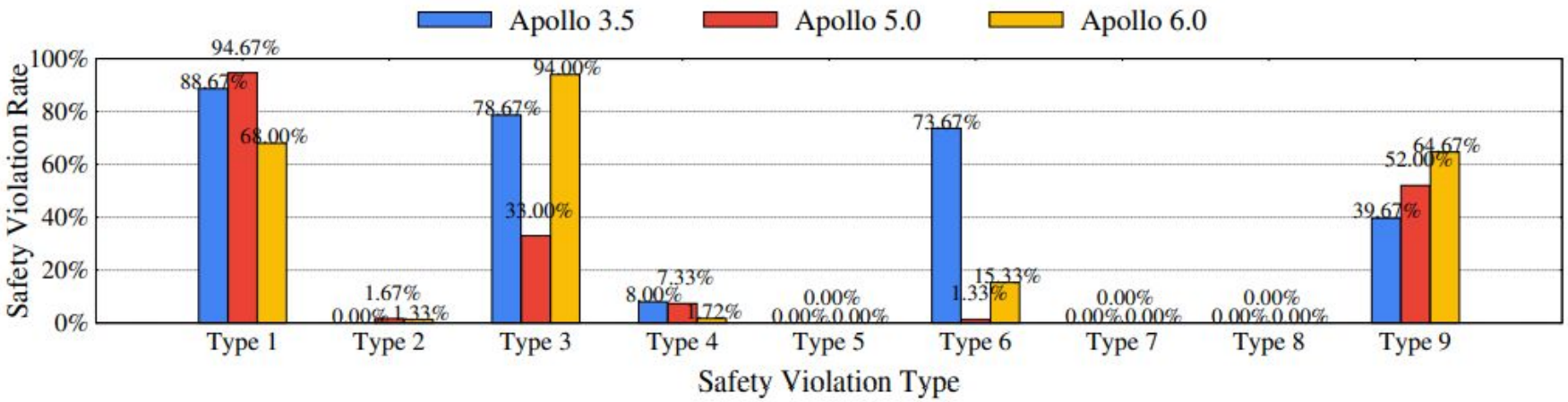
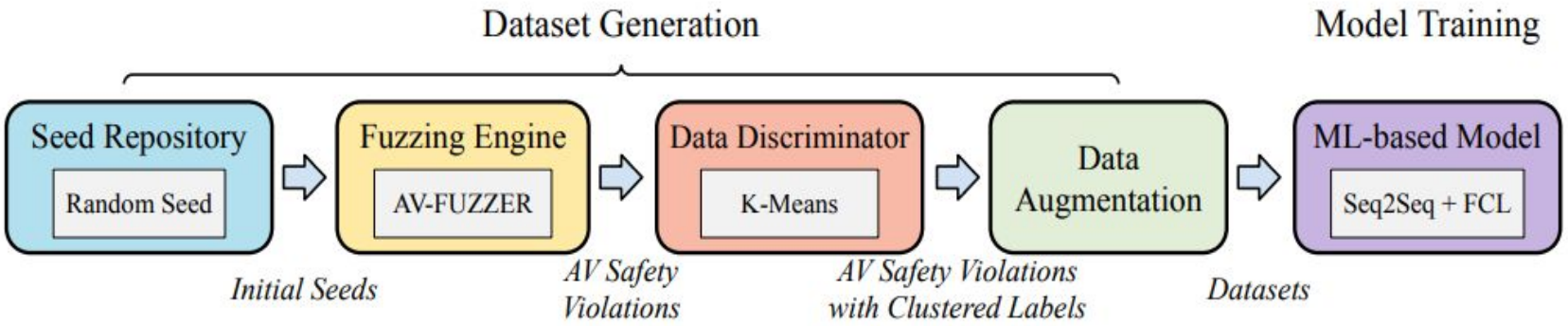


Fig. 5: AV safety violation rates among 9 accident types under three different ADS versions.

Takeaways

- Different ADSs show different vulnerabilities under different safety violation types
- Human driver accidents cannot fully show vulnerability of ADS

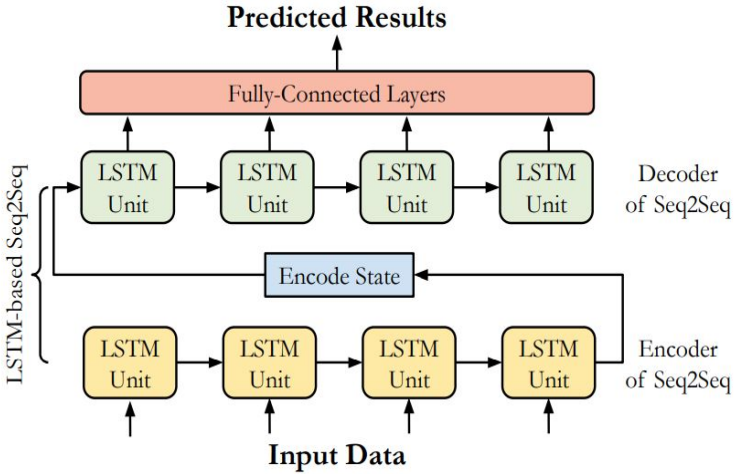
Framework Design : Model Creation



ML-Based Model

Input: Past 1.25 second's vehicle running data

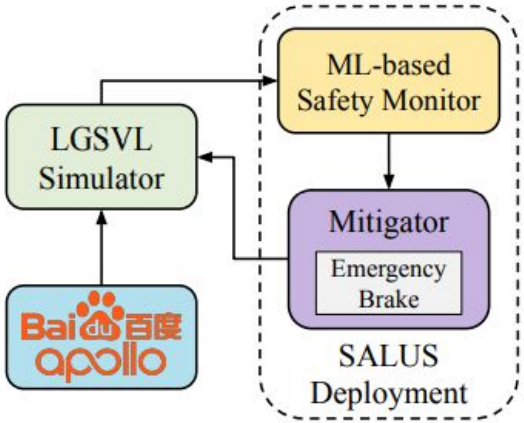
- NPC and AV's current speed
- Euclidean distance of NPC and AV's location
- Euclidean distance of NPC and AV's steering angle
- Indicator that judge "is NPC in front of AV?"
- ……



Output: Safety indicators for next 2 seconds

Framework Design : Model Deployment

- How does the model deployed?
- How does safety monitor work?



Experimental Setup

Machine Specification:

- Ubuntu 18.04 machine with 32GB RAM
- AMD 5900X CPU (12-core/24- thread)
- NVIDIA GTX 1080 Ti GPU card

Environment setup

- One npc
- 2-lane straight road

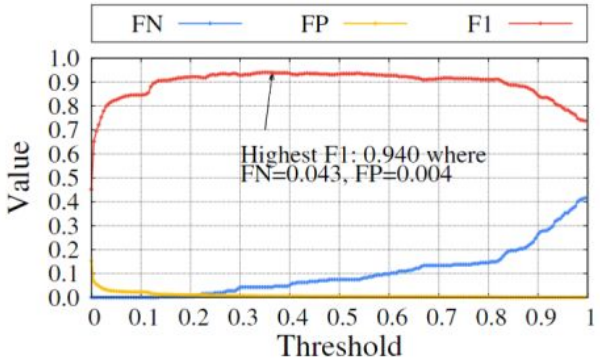
Evaluation: Metrics

- **False Negative:** (number of FN time slots) / (number of total time slots)
- **False Positive:** (number of FP time slots) / (number of total time slots)
- **F1 Score:** $$\mathrm{F1} = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})}$$
- **Safety Violation Rate:** (number of safty violations trials) / (number of total trials)

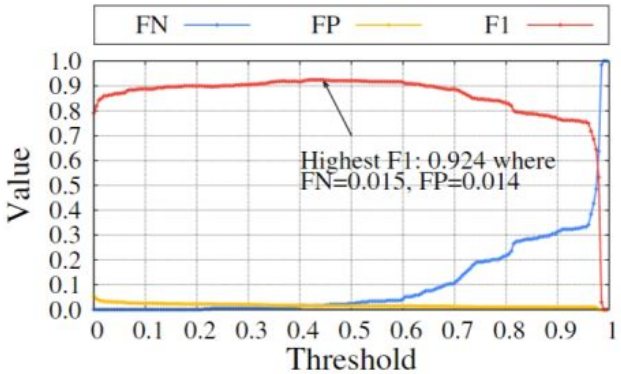
Evaluation: ML Model Prediction Accuracy

- Test dataset for each group in each ADS:
  - 100 trials will lead to AV safety violation
  - 100 trials will **NOT** lead to AV safety violation
  - 50 totally random trials not lead to AV safety violation

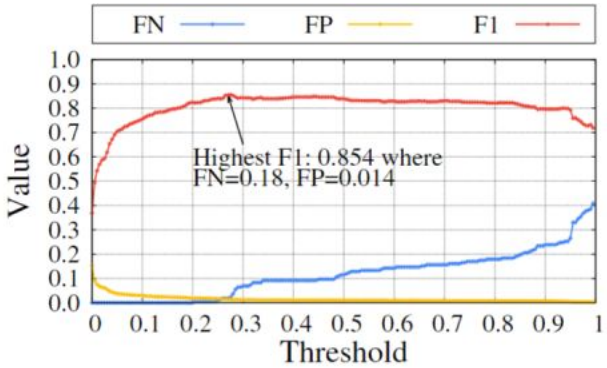
Average F1 score is 0.889



Apollo 3.5 Group 1



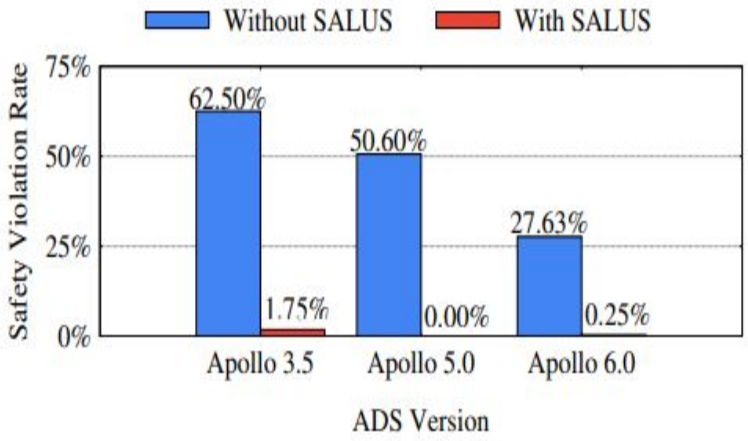
Apollo 5.0 Group 2

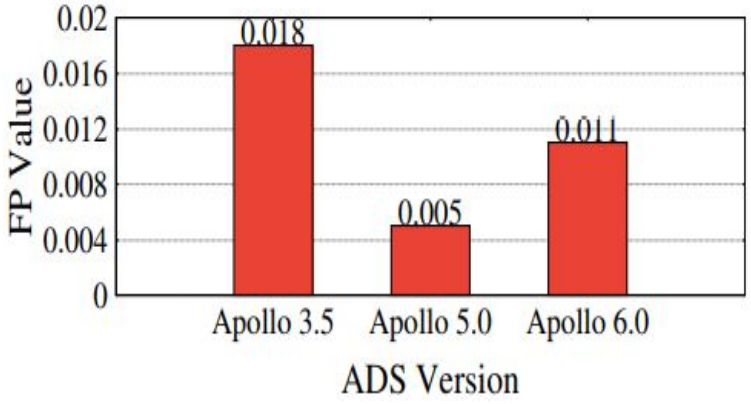


Apollo 6.0 Group 4

Evaluation: Safety Violation Mitigation Effiency

- Test dataset for each ADS
  - 200 randomly generated safety violation trials for each group in each ADS
  - 200 totally random trials for testing the FP value for each ADS
- SALUS can mitigate more than 97.2% of AV safety violations compared with an ADS without any protection and FP value is less than 0.018





Evaluation: Cross-Validation of Models and ADS Versions

ADS version used their own models perform better

TABLE II: The cross validation of safety violation rates across different models and ADS versions.

	Apollo 3.5	Apollo 5.0	Apollo 6.0
Model for Apollo 3.5	**1.75%**	6.24%	7.18%
Model for Apollo 5.0	0.00%	**0.00%**	5.55%
Model for Apollo 6.0	0.75%	0.60%	**0.25%**

TABLE III: The cross validation of FP values across different models and ADS versions.

	Apollo 3.5	Apollo 5.0	Apollo 6.0
Model for Apollo 3.5	**0.018**	0.069	0.072
Model for Apollo 5.0	0.027	**0.005**	0.056
Model for Apollo 6.0	0.086	0.060	**0.011**

Conclusion & Future Work

- Our goal is to eliminate AV safety violations
  - Inital study shows different ADSs exhibit very different sensitivity to different safety violations
  - Proposed a framework for SALUS that able to Creation and Deployment safety monitor
    - NPC’s trajectoris which will lead the AV safety violations fall into identifiable patterns
    - ML-based prediction model
    - Enable the maxium AV-safety

- Future work
  - Accelerating the fuzzing engine in SALUS creating workflow
  - Updating the mitigation strategy
  - Test more scenario