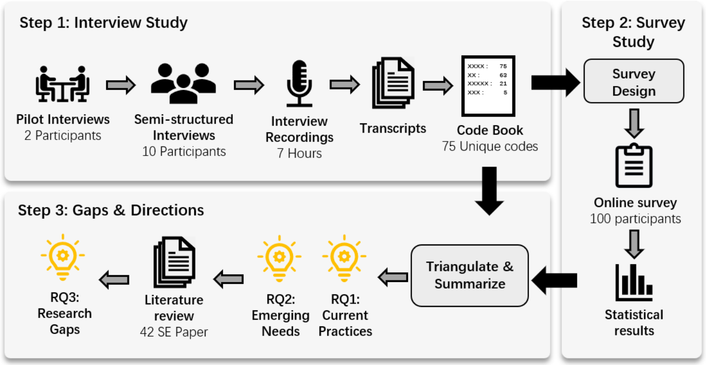
Towards Robust Autonomous driving Systems

Dr. James Xi Zheng

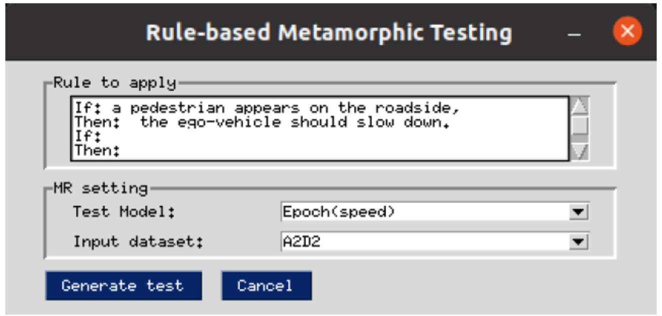
Talk Agenda

A Study



Test Reduction

Test Generation



Future/Ongoing Work

Autonomous Driving Landscape

Baidu's Apollo Offers fully driverless ride-hailing service in Wuhan and Chongqing [1]

Multiple Autonomous Driving Companies claim having 20+M miles on public roads (Baidu, Waymo) while Zoox and Cruise are gaining more licenses to expand public road deployment

The global autonomous vehicle market was valued at $76,126.43 million in 2020, and is projected to reach $2,161,787.29 million by 2030 [2]

[1] https://www.autonomousvehicleinternational.com/news/robotaxis/baidu-apollo-day-expanded-robotaxi-operations-and-new-autonomous-driving-software-and-hardware.html
[2] https://www.alliedmarketresearch.com/autonomous-vehicle-market

Autonomous Vehicle Levels

- Created by SAE International and adopted by the U.S. Department of Transportation (DoT) and the National Highway Traffic Safety Administration (NHTSA) in November 2017.

- LEVEL 0 – majority of the vehicles
- LEVEL 1 - steering, throttle, braking (a set of operations referred to as ST&B) (Cruise control)
- LEVEL 2 – more ST&B but more permanent way (At this level, the car can act autonomously, but requires the driver to monitor the driving at all times and be prepared to take control at a moment's notice.)
- **LEVEL 3** - the car can monitor its surroundings and handle ST&B but in specific condition (highway)
- **LEVEL 4** – fully automated but needs driver in specific condition (weather conditions limits)
- LEVEL 5 – DRIVERS GONE, only passenger. No steering wheel

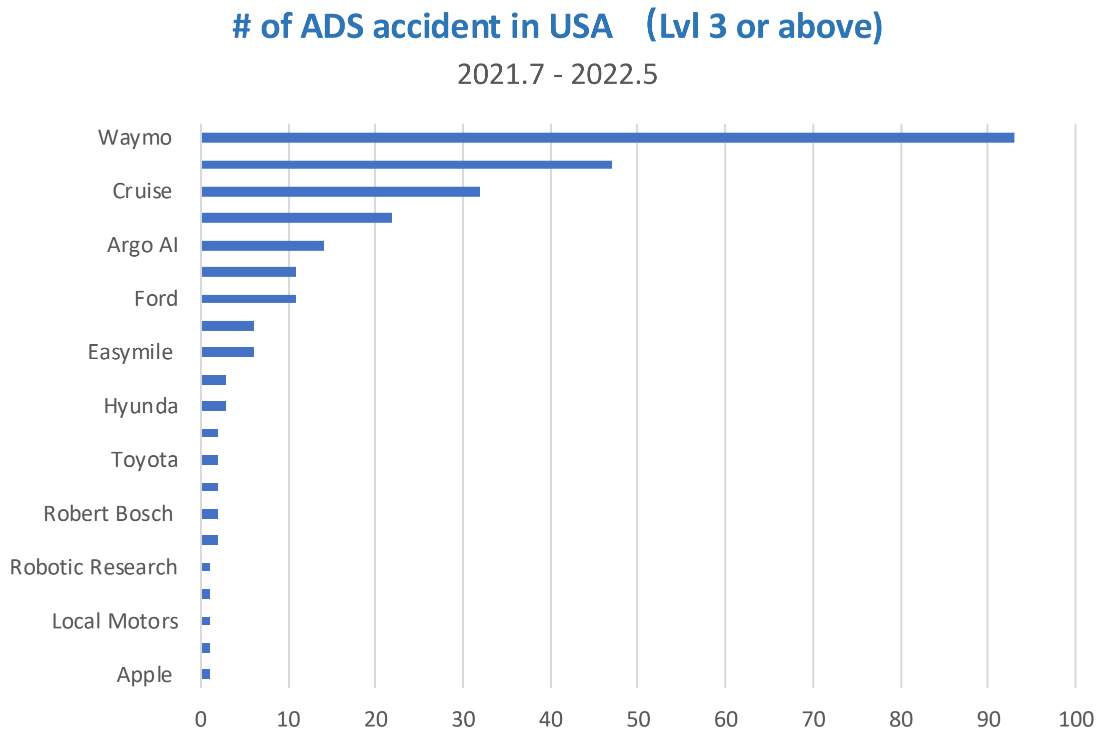
Autonomous Driving Systems (ADS) Safety Issues

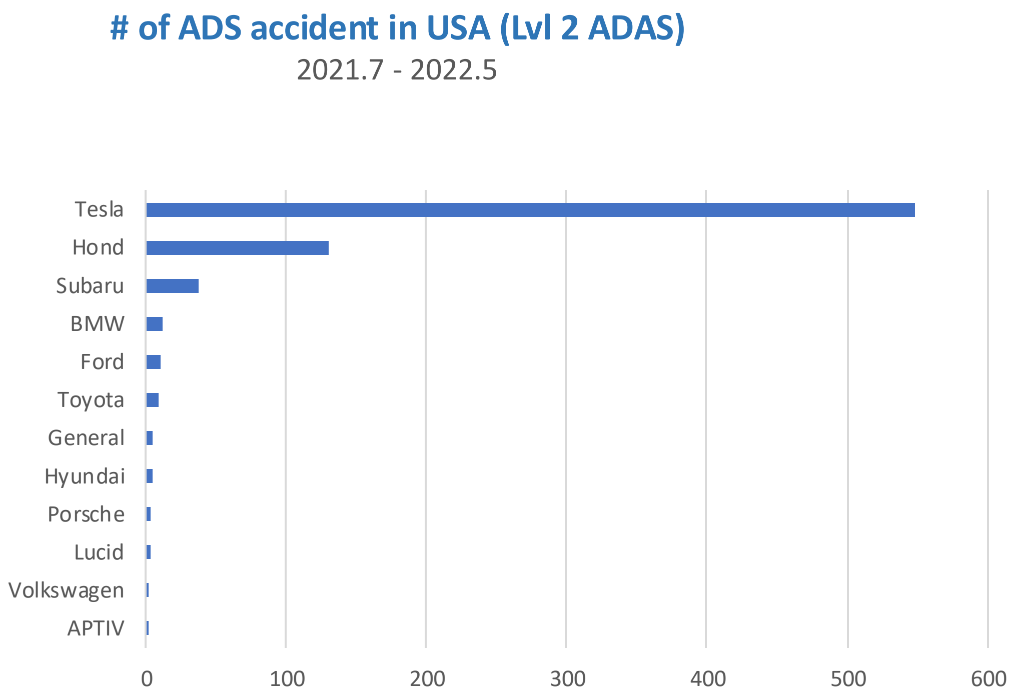
Fig 1. Tesla was already under investigation for a series of accidents involving the company's autopilot and emergency services vehicles.

Laguna Beach Police Department/AP

Fig 2. Xiaopeng P7 crashed into stationary turnover vehicle.
Image from
https://baijiahao.baidu.com/s?id=1740929278376223344&wfr=spider&for=pc

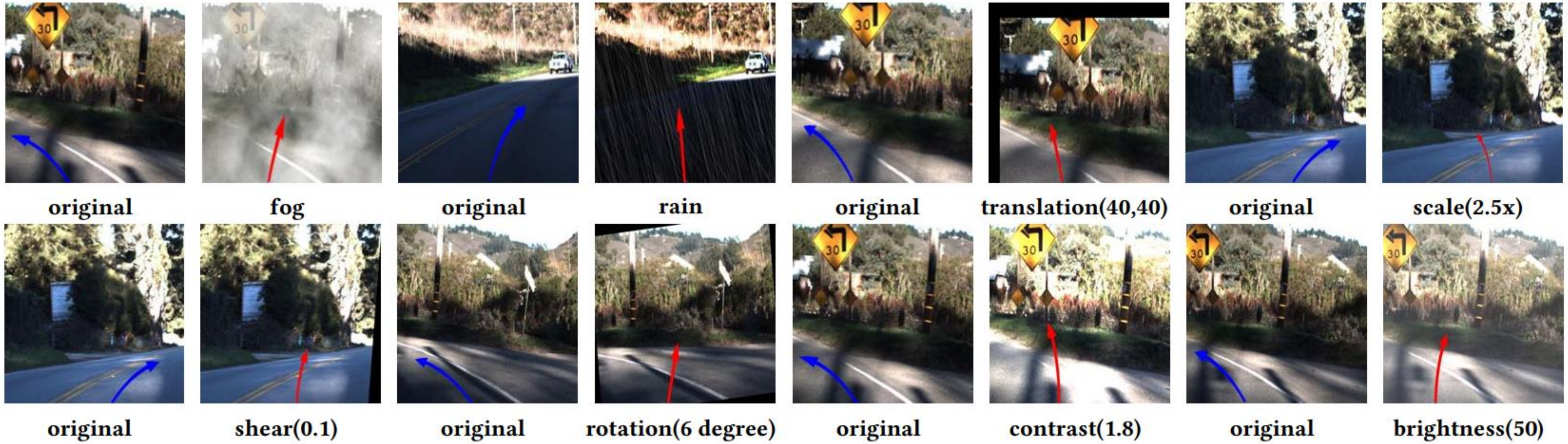
Autonomous Driving Systems (ADS) Safety Issues





Data from: https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting#data

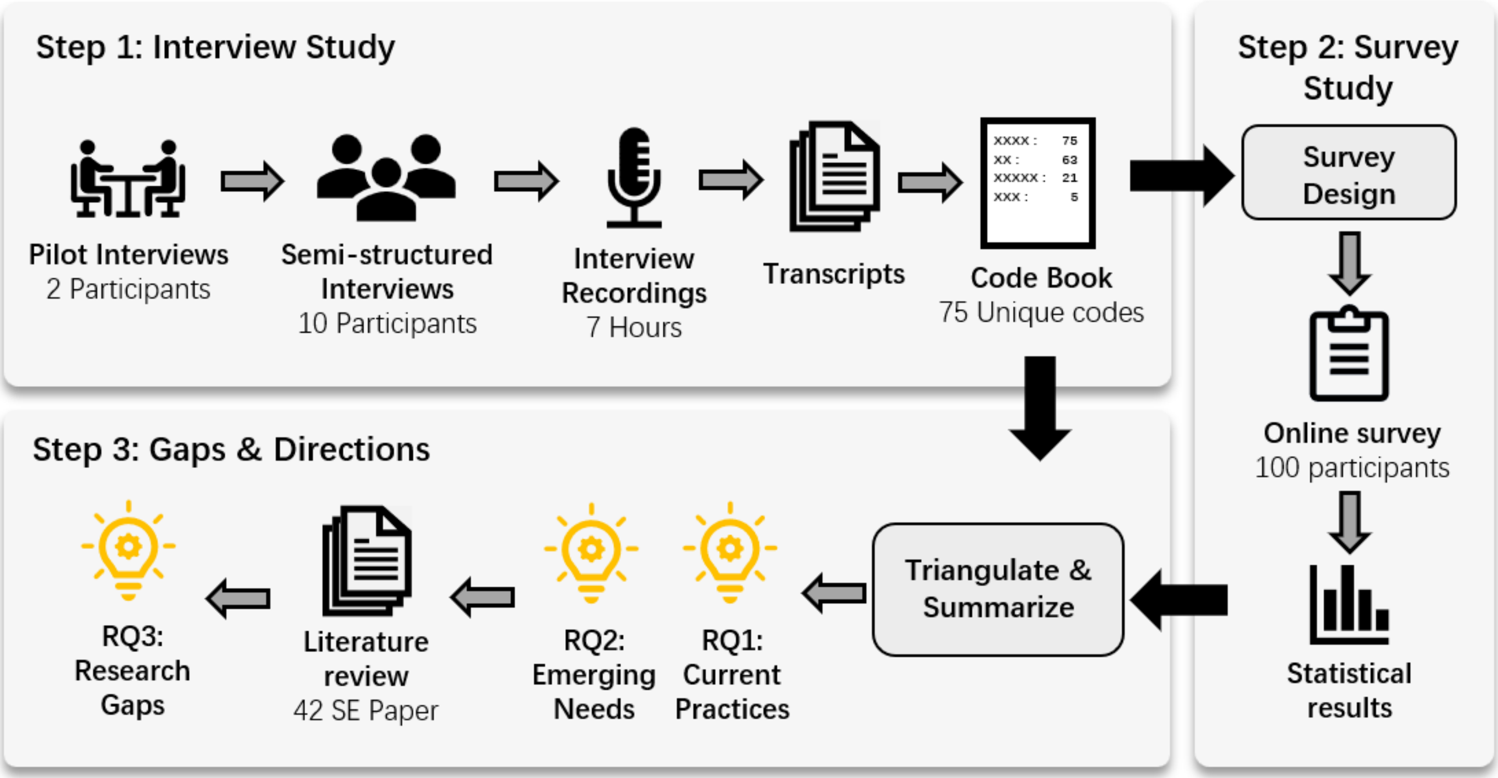
ADS Testing Research – A SOTA work (ICSE'18 1000+ citations)



Sample images showing erroneous behaviors detected by DeepTest[1] using synthetic images.
For original images the arrows are marked in blue, while for the synthetic images they are marked in red [1]

[1] Tian, Yuchi, et al. "Deeptest: Automated testing of deep-neural-network-driven autonomous cars." Proceedings of the 40th international conference on software engineering. 2018.

Testing of Autonomous Driving Systems: -where are we and where should we go? (FSE’22)



Research Methodology

[1] G Lou, Y Deng, X Zheng, M Zhang, T Zhang. "Testing of Autonomous Driving Systems: Where are we and where should we go?" The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)

RQ1

COMMON PRACTICES

1. The majority of ADS practitioners reported to work on multi-module ADSs rather than end-to-end driving models.
2. ADS developers also need to construct segments of driving recordings to test DL models, which take multi-modal sensor data as input, not just road images.
3. ADS practitioners design various driving scenarios based on real-world driving scenes, public benchmarks, traffic laws and regulations, and crash reports to test an ADS in the field.

⋮

RQ1

COMMON PRACTICES

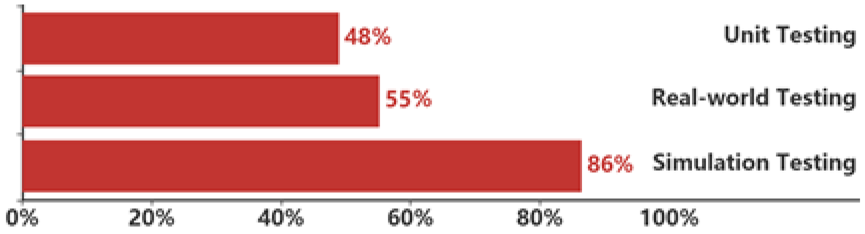


Fig 7. Testing methods.

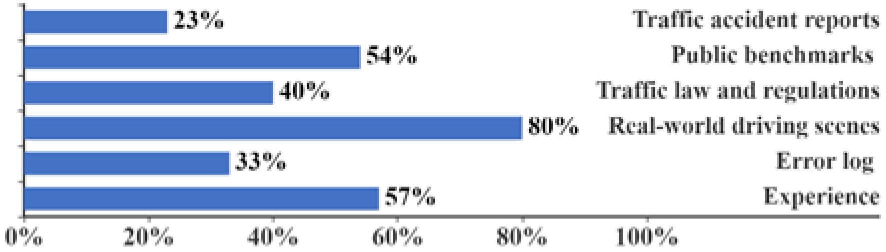


Fig 8. Sources of driving scenarios.

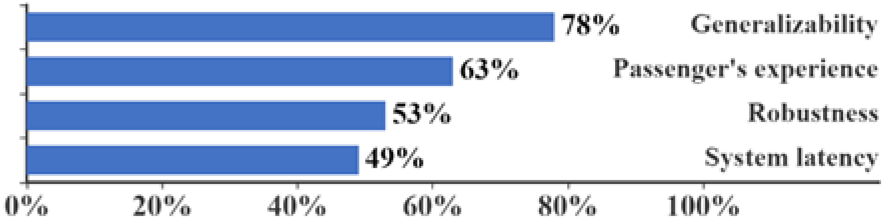


Fig 9. System-level metrics in real-world testing.

RQ2

4 EMERGENT NEEDS

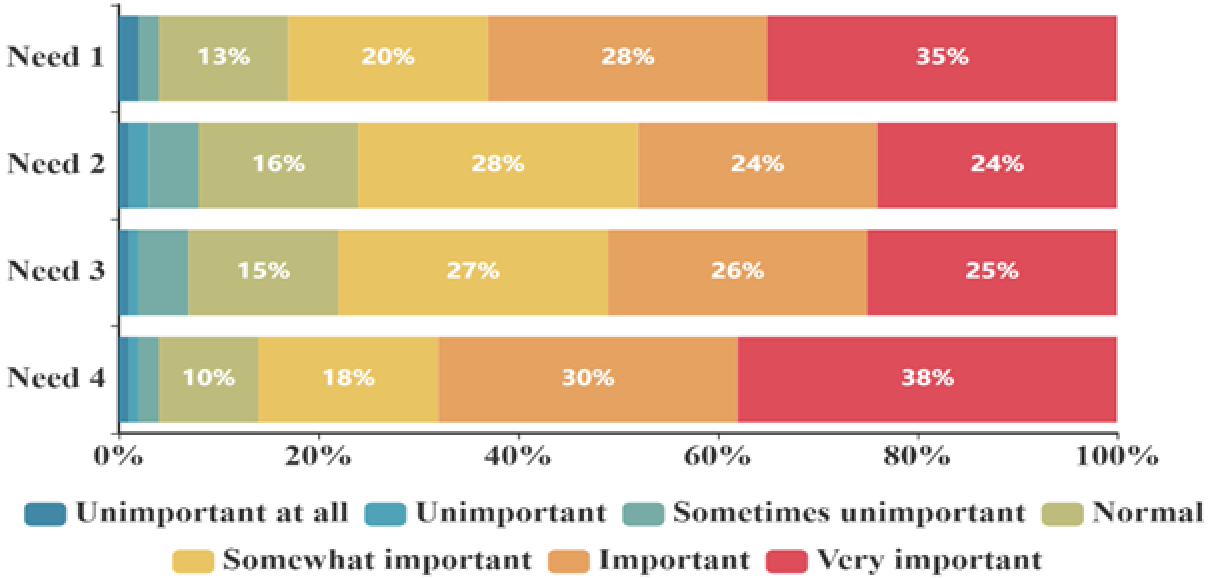


Fig 11. The importance of four common needs voted by survey participants.

Emergent Needs:

1. Identifying possible corner cases
2. Speeding up ADS testing
3. Tool support for constructing complex driving scenarios
4. Tool support for data labeling

RQ3

GAP AND FUTURE DIRECTION

Need 1: Identifying possible corner cases and unexpected driving scenarios

Fig 12. Finding corner cases is a long tail problem

Current Situations:

- Finding corner cases is a long tail problem
- Large companies find corner cases leveraging large-scale testing
- Small companies cannot afford it

RQ3

GAP AND FUTURE DIRECTION

Need 1: Identifying possible corner cases and unexpected driving scenarios

Fig 13. An example of complex driving scenario.

SE Solution:

- Knowledge-based methods
- Search-based methods

Gaps:

- Simple cases in existing works
- Large search space

Future Directions:

- Investigate on transformations in complex driving cases (e.g., lane merging, overtaking)
- Generate fitness functions from existing driving requirements
- Identify corner cases on multi-modal sensor

RQ3

GAP AND FUTURE DIRECTION

Need 2: Speeding up ADS testing

At least 10^7-hour testing (about 1141 vehicle-years) is required to verify that the failure rate falls within the requirement.

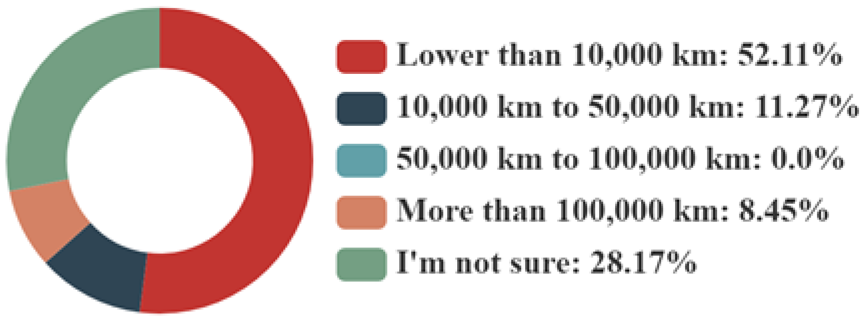


Fig 13. Length of on-road testing performed by industrial survey participants

Current Situation:

- Long test (*about 1141 vehicle-years*) is required
- Small companies struggle to support large fleets
- Simulation platform with speed up functions is applied, but still not enough

RQ3

GAP AND FUTURE DIRECTION

Need 2: Speeding up ADS testing

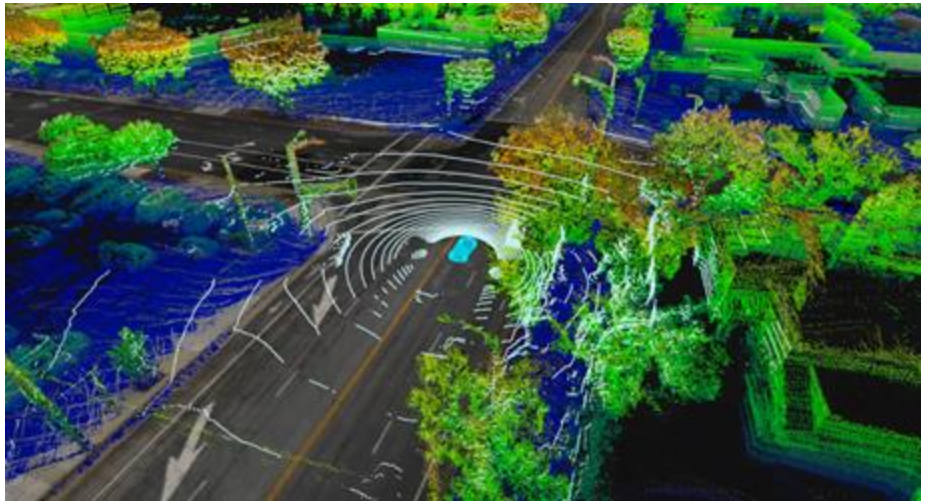


Fig 14. Rich informations in HD maps

SE Solutions:

- Test Selection and Test Prioritization

Gaps:

- Simple metrics (Summarize)
- Cannot handle multi-module structure

Future Directions:

- Use HD map with rich information as the target of metrics.
- Use multi-objective optimization to combine for inner model behaviors metrics and logic-based test metrics.

RQ3

GAP AND FUTURE DIRECTION

Need 3: Tool support for constructing complex driving scenarios

Fig 15. Apollo Dreamview Simulator

Current situation:

- It is cumbersome to use existing simulation platforms to construct a test case.
- Existing simulation platforms only provide low-level APIs and domain-specific languages (DSLs) to construct driving scenarios.

RQ3

GAP AND FUTURE DIRECTION

Need 3: Tool support for constructing complex driving scenarios

Fig 15. Apollo Dreamview Simulator

SE Solution:

- AC3R[1] (Limited Scenarios)
- ADSML[2] (Focus on formal verification)

Future directions:

- Automatically extracting driving scenarios from more diverse resources using NLP and CV techniques
- Semi-automatically generating batches of similar driving scenarios using high level DSL.

[1] Alessio Gambi, Tri Huynh, and Gordon Fraser. Generating effective test cases for self-driving cars from police reports. In Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pages 257–267, 2019.

[2] Dehui Du, Jiena Chen, Mingzhuo Zhang, and Mingjun Ma. Towards verified safety-critical autonomous driving scenario with adsml. In 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), pages 1333–1338. IEEE, 2021.

RQ3

GAP AND FUTURE DIRECTION

Need 4: Tool support for data labeling

Fig 16. ADS labels require high precision

Current Situation:

- Massive amount of driving data
- Although assistive labeling tools exists, labeling driving data still requires many manual effort
- Quality of Manually labeled data is still a concern

RQ3

GAP AND FUTURE DIRECTION

Need 4: Tool support for data labeling

Fig 19. Semi-labeling provided by Segments.ai

SE Solution:

- Using surprise adequacy to avoid labeling similar inputs.

Gaps:

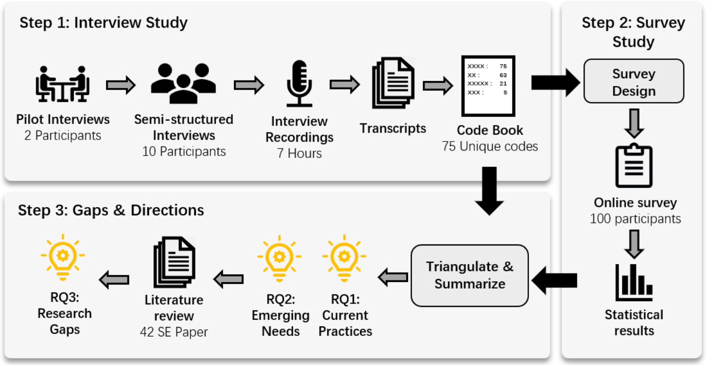
- Still rely on human laboring

Future directions:

- Automatic/semi-automatic labeling tools
- Developing input validation or outlier detection for data cleaning

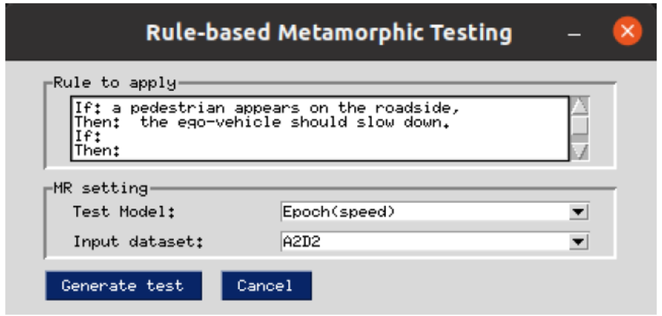
Talk Agenda

A Study



Test Reduction

Test Generation



Future/Ongoing Work

Scenario-based Test Reduction and Prioritization for Multi-Module Autonomous Driving Systems (STRAP) (FSE'22)

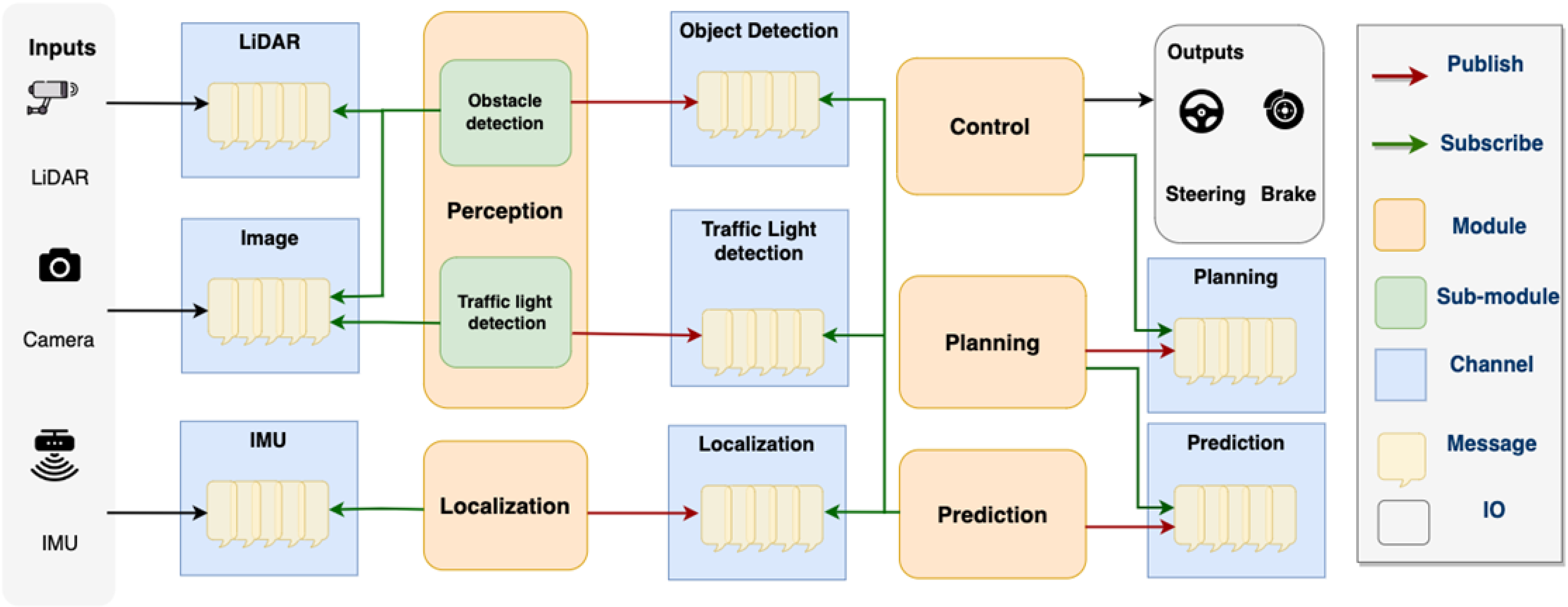
- Driving recordings contain redundant driving scenarios
- Speed up (regression) testing of autonomous driving systems by test reduction and prioritization
- Traditional methods failed

Redundant driving scenarios on highway

[1] Y Deng, X Zheng, MS Zhang, G Lou, T Zhang. *"Scenario-based Test Reduction and Prioritization for Mutli-Module Autonomous Driving Systems" The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*

ADS Testing

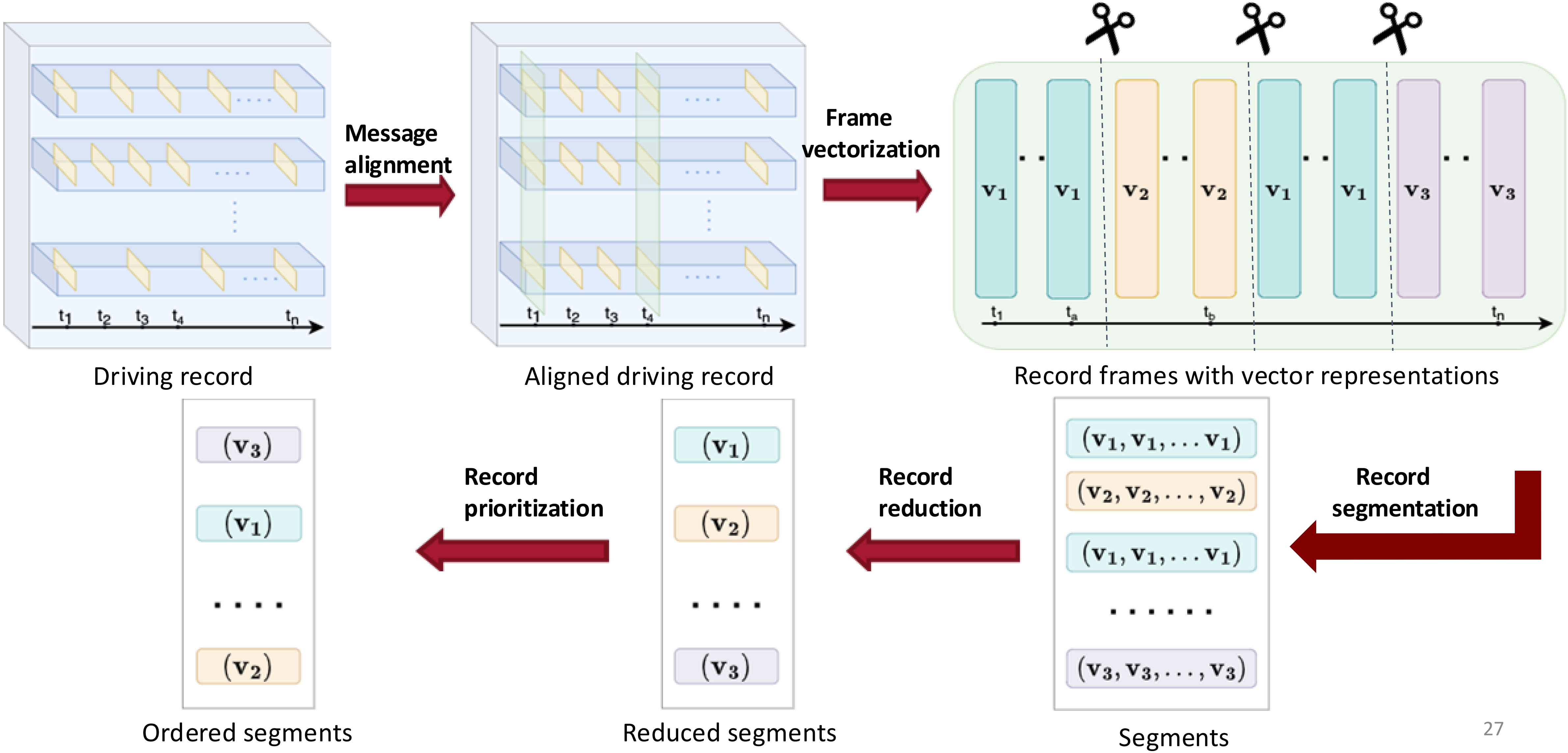
Multi-module Autonomous Driving System (ADS)



The architecture and data flow of a multi-module ADS

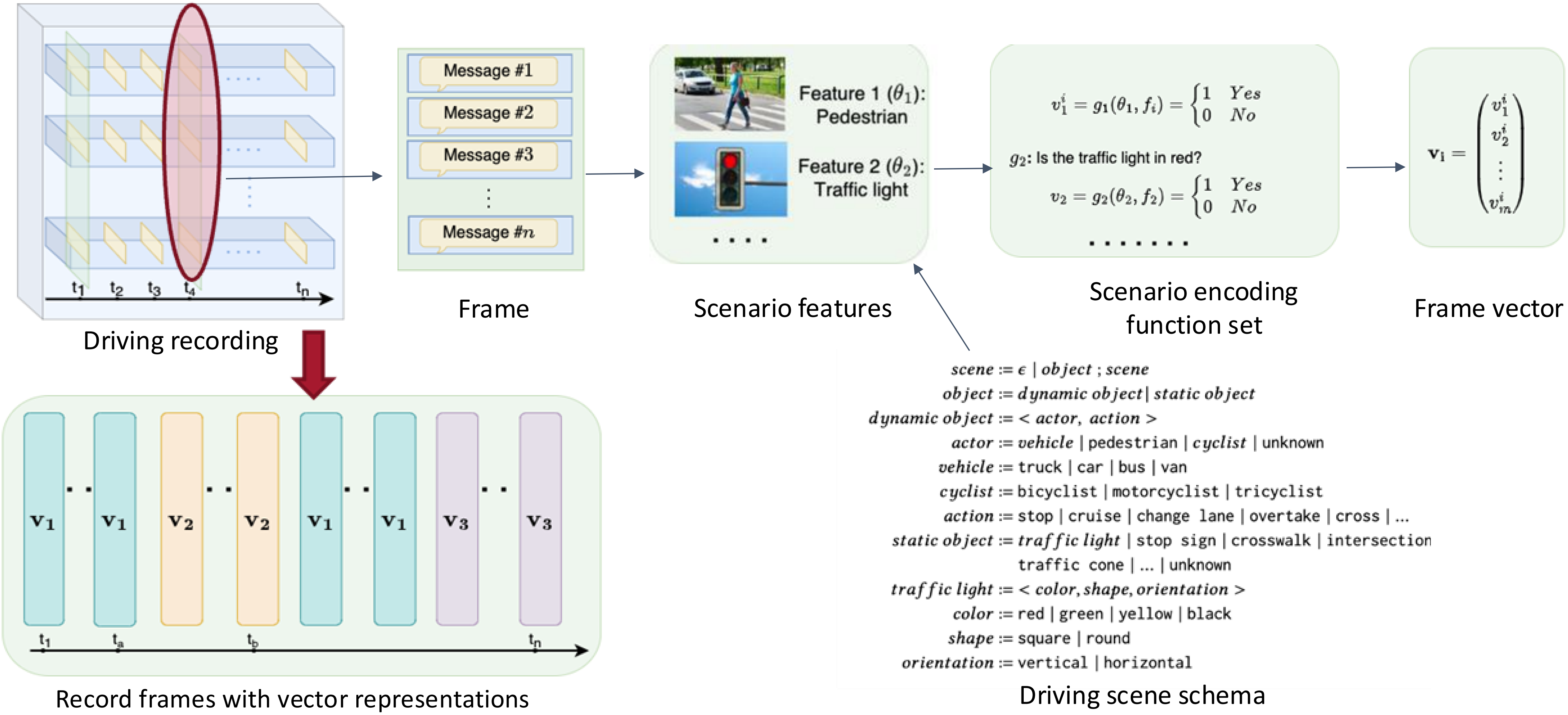
Methodology

Overview



Methodology

Frame vectorization



Experiments

Research Questions and Evaluation Metrics

- **RQ1:** To what extent can the proposed test reduction method reduce a given driving recording?
    - **Reduction ratio of testing time**
- **RQ2:** Compared with an original recording, how effective is the reduced recording in detecting faults?
    - **Fault coverage**
- **RQ3:** Compared with alternative prioritization methods, how effective is the proposed test prioritization method in detecting faults?
    - **Average percentage of faults detected (APFD)**: measures the capability of fault detection per percentage of test cases execution
    - **TOP-K**: measures the amount of segments required for finding the first fault

Experiments

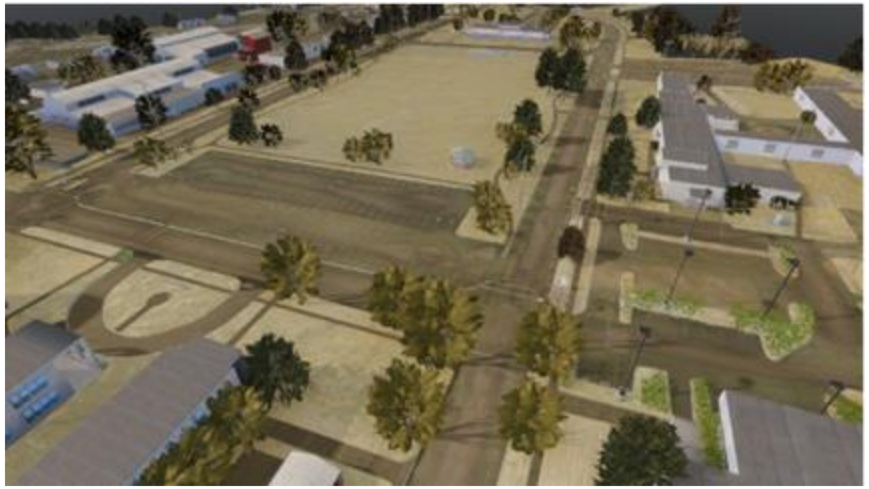
Baselines for RQ3

- Proposed test prioritization methods:
  - Prioritize segments based on the semantic coverage (SC)
  - Prioritize segments based on the rarity and semantic coverage (RSC)
- Test prioritization baselines:
  - Prioritize segments in chronological order (CH)
  - Prioritize segments randomly (RD)
  - Prioritize segments by the number of calls on the changed function (CC)

Experiments

Test Suite Creation

- Experiment simulator: LGSVL
- ADS under testing: Baidu Apollo 5.0
- Collect driving recordings from three maps: Cubetown, Gomentum, and Shalun
- Faults are injected in ADS modules: Traffic light detection; Obstacle detection; Prediction; Planning
- For each module, 9 modified versions are created

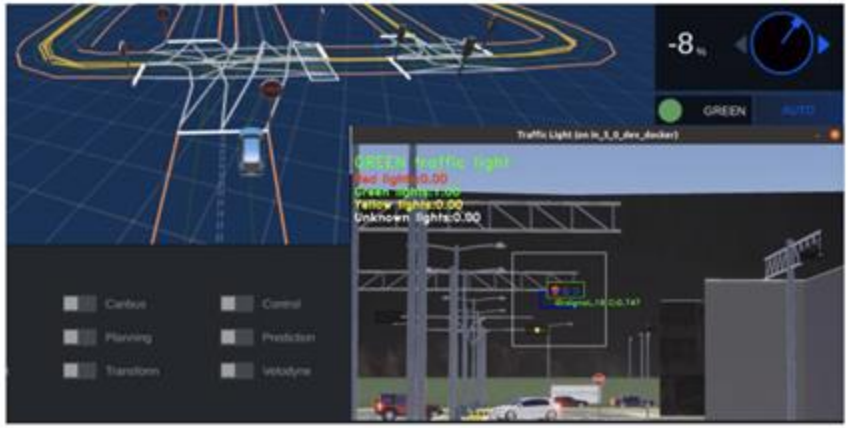


```
- DEFINE_double(buffer_out_routing, 2.0,
+ DEFINE_double(buffer_out_routing, -7.0,
              "buffer for select out lane for boundary");
  // planning trajectory output time density control
  DEFINE_double(
```

(a) A real fault of correcting buffer size

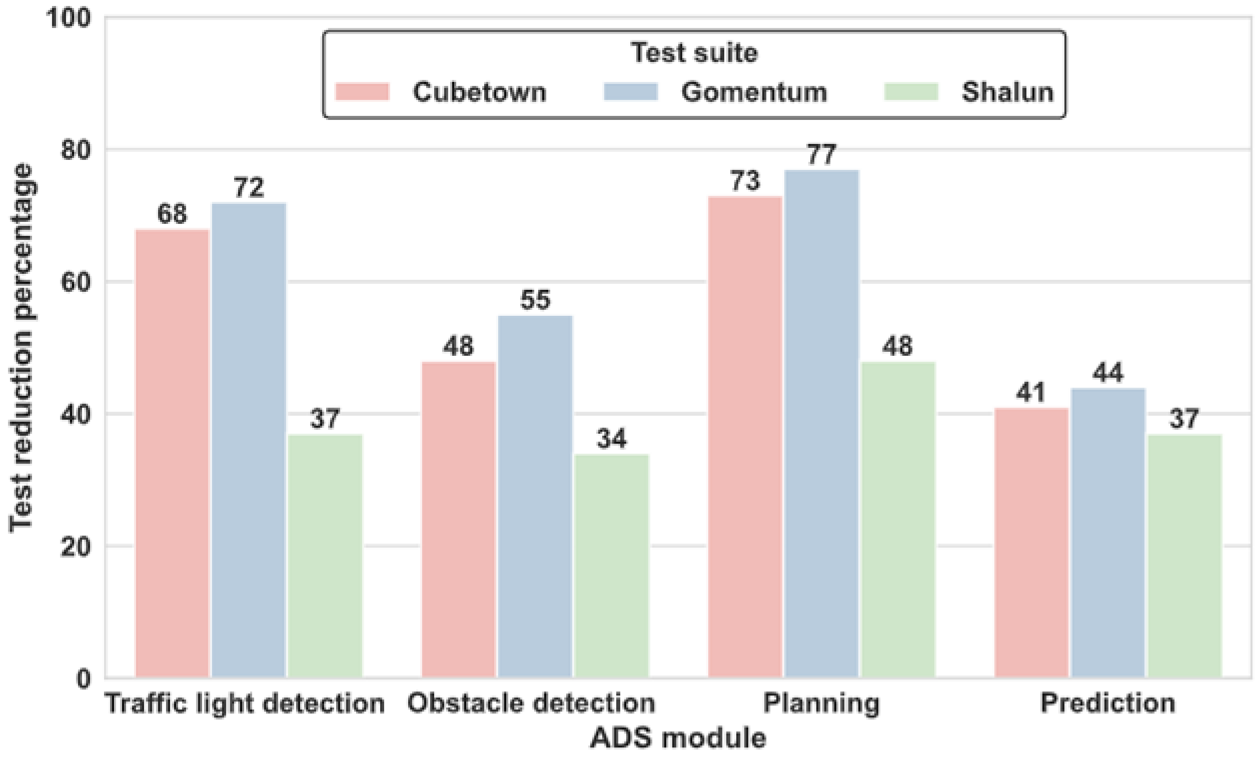
```
    point.z = pt.z();
    point.intensity = static_cast<float>(pt.intensity());
    frame->cloud->push_back(point, static_cast<double>(pt.
      /* original code was : FLT_MAX, i, 0); */
                        FLT_MAX, i, -1);
  }
```

(b) A mutant of changing constant value



Results

Effectiveness of Test Reduction



Test reduction ratio (RQ1)

- The reduction ratio of testing time achieves 77% at most for the planning module in the test suite Gomentum.
- The test reduction ratios are relatively lower for modules in Shalun, which is caused by the complexity of the map.

Results

Effectiveness of Test Reduction

Test suite	ADS module	Total faults	Covered faults
Cubetwon Gomentum Shalun **Total**	Traffic light detection	13 19 78 **109**	13 (100%) 19 (100%) 72 (92.3%) **104 (95.4%)**
Cubetwon Gomentum Shalun **Total**	Obstacle detection	49 28 27 **94**	49 (100%) 28 (100%) 27 (100%) **94 (100%)**
Cubetwon Gomentum Shalun **Total**	Planning	42 10 64 **116**	42 (100%) 10 (100%) 64 (100%) **116 (100%)**
Cubetwon Gomentum Shalun **Total**	Prediction	34 61 19 **114**	34 (100%) 61 (100%) 19 (100%) **114 (100%)**
Total	**Total**	**433**	**428 (98.8%)**

Fault coverage (RQ2)

- Almost all faults in the original driving recordings can be covered by reduced driving segments
- The missed faults in Shalun for the traffic light detection module may be caused by short-duration glitch of the detection module on vertical traffic lights in Shalun.

Results

Effectiveness of Test Prioritization

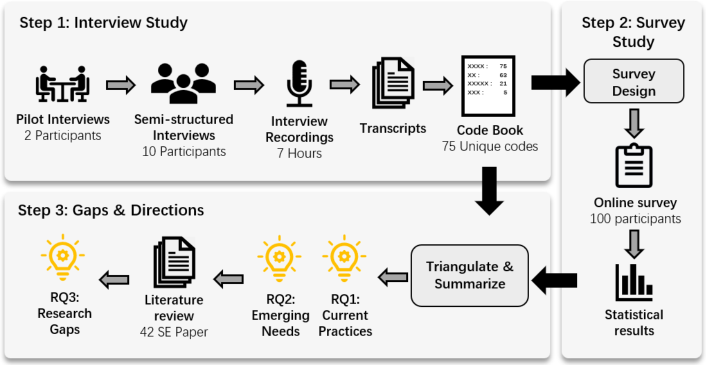
Test suite	ADS module	Top K CH	Top K RD	Top K CC	Top K SC	Top K RSC	APFD CH	APFD RD	APFD CC	APFD SC	APFD RSC
Cubetwon	Traffic light detection	6.00	4.26	5.50	5.00	**2.00**	0.43	0.50	0.39	0.50	**0.67**
Gomentum		2.00	3.57	6.14	**1.00**	**1.00**	0.50	0.50	0.25	**0.82**	0.76
Shalun		1.83	1.46	2.17	2.00	**1.33**	0.56	0.50	0.51	**0.56**	0.54
Total		3.28	3.10	4.60	2.67	**1.44**	0.50	0.50	0.38	0.63	**0.66**
Cubetwon	Obstacle detection	1.89	2.47	2.44	3.00	3.00	0.56	0.50	0.53	0.59	**0.59**
Gomentum		3.67	1.74	2.22	1.00	1.00	0.51	0.50	0.34	**0.44**	0.47
Shalun		2.00	1.17	1.22	1.00	1.00	0.48	0.50	0.50	**0.53**	0.53
Total		2.52	1.79	1.96	1.67	**1.67**	0.52	0.50	0.46	0.52	**0.53**
Cubetwon	Planning	3.00	1.38	1.00	1.00	1.00	0.37	0.50	0.58	0.59	**0.59**
Gomentum		7.00	4.48	7.14	7.00	2.00	0.19	0.50	0.17	**0.19**	0.81
Shalun		3.00	1.60	1.00	1.00	1.44	0.49	0.50	0.56	**0.53**	0.58
Total		4.33	2.49	3.05	3.00	**1.48**	0.35	0.50	0.44	0.44	**0.66**
Cubetwon	Prediction	4.44	3.27	4.44	2.22	2.22	0.39	0.50	0.48	0.61	**0.62**
Gomentum		1.22	1.13	1.11	1.00	1.00	0.50	0.50	0.50	**0.50**	0.51
Shalun		4.63	5.82	6.50	3.75	2.00	0.55	0.50	0.45	**0.61**	0.64
Total		3.43	3.41	4.02	2.32	**1.74**	0.48	0.50	0.48	0.57	**0.59**
Total	**Total**	3.39	2.70	3.41	2.42	**1.58**	0.46	0.50	0.44	0.54	**0.61**

Effectiveness of Test Prioritization (RQ3)

- The proposed methods SC and *RSC* outperforms other baselines.
- Empirically, the correlation between semantic rareness and faults is very positive.

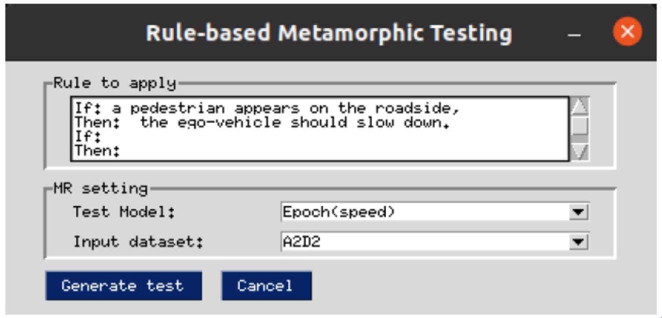
Talk Agenda

A Study



Test Reduction

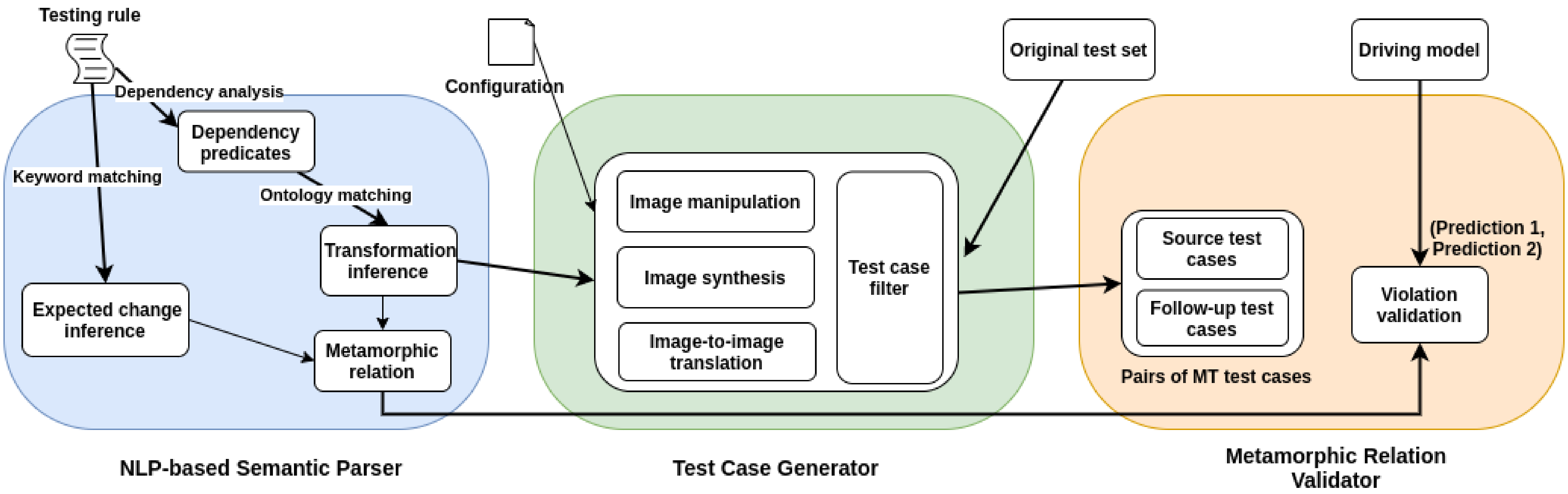
Test Generation



Future/Ongoing Work

A Declarative Metamorphic Testing Framework for Autonomous Driving (TSE'23)

OVERVIEW OF RMT



Architecture of RMT

[1] Deng, Y., Zheng, X., Zhang, T., Liu, H., Lou, G., Kim, M. and Chen, T., 2023. A Declarative Metamorphic Testing Framework for Autonomous Driving. IEEE Transactions on Software Engineering, to-appear

Methodology

TESTING RULE SPECIFICATION

A testing rule is :

- the description of a testing scenario and the expected change of the driving behavior written in natural language following IFTTT (If-This-Then-That) paradigm.
- derived from traffic rules.

Country/District	Traffic rules
NSW, Australia [1]	When you see potential hazards, slow down and prepare to stop, for example when pedestrians are close to the road or when other vehicles may turn in front of you.
California, USA [2]	A 3-sided red YIELD sign indicates that you must slow down and be ready to stop, if necessary, to let any vehicle, bicyclist, or pedestrian pass before you proceed.
Germany [3]	Drive more slowly at night because you cannot see as far ahead and you will have less time to stop for a hazard.

Examples of Traffic rules

→

If: a pedestrian appears on the roadside,
Then: the ego-vehicle should slow down.

If: a speed limit sign appears on the roadside,
Then: the ego-vehicle should slow down.

If: the driving time changes into night,
Then: the ego-vehicle should slow down.

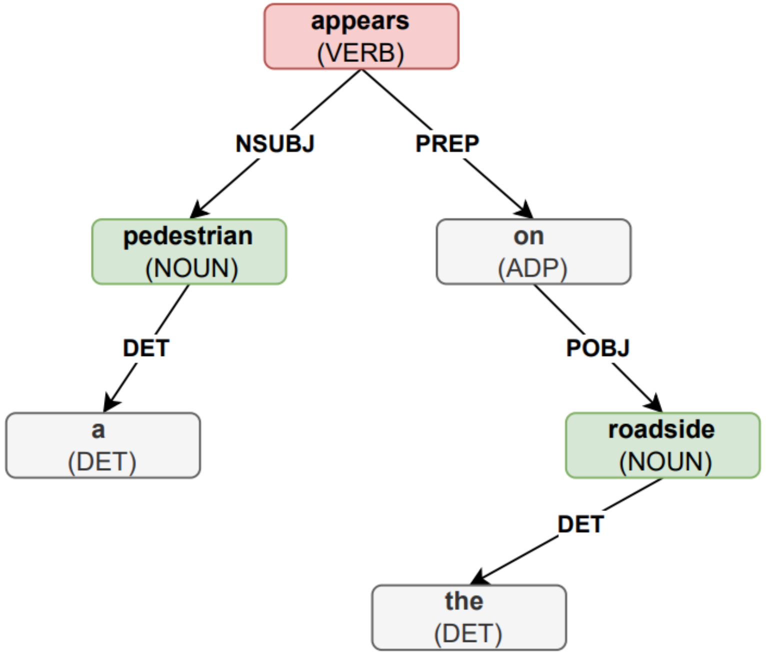
Testing rules following IFTTT syntax

Methodology

NLP-BASED SEMANTIC RULE PARSING

1. Dependency analysis

If a pedestrian appears on the roadside



Dependency parsing

Methodology

NLP-BASED SEMANTIC RULE PARSING

2. Ontology-based Information Extraction

Category	Level-1 Subcategory	Level-2 Subcategory	Properties
Road network	Road part	Lane	direction: [forward, reverse] orientation: [vertical, horizontal] position: [left, right]
		Line	Type: [solid, dash] Color: [white, yellow]
		Crosswalk	Orientation: [vertical, horizontal]
		Sidewalk	—
	Traffic infrastructure	Traffic sign	Type: [stop, speed limit, turn] Shape: [circle, square]
		Traffic light	Color: [red, yellow, green]
		Stop and yield line	—
Object	Static object	Tree	—
		Building	—
	Dynamic object	Pedestrian	—
		Vehicle	Type: [car, truck, van, school bus] Color: [white, black, blue, ...]
		Bicyclist	—
Environment	Weather	rainy, cloudy, snowy	Level: [light, normal, heavy]
	Time	day, night	—

Driving scene ontology

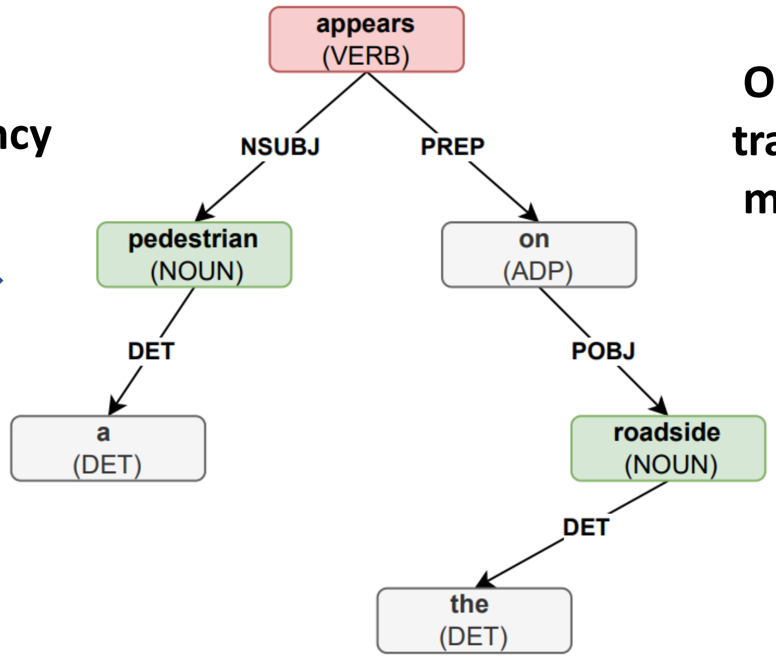
Methodology

NLP-BASED SEMANTIC RULE PARSING

2. Ontology-based Information Extraction

If a pedestrian appears on the roadside

Dependency parsing



Ontology & transformation matching

Ontology(Pedestrian)
Ontology(Roadside)
Transformation(Add)

Methodology

METAMORPHIC TEST GENERATOR

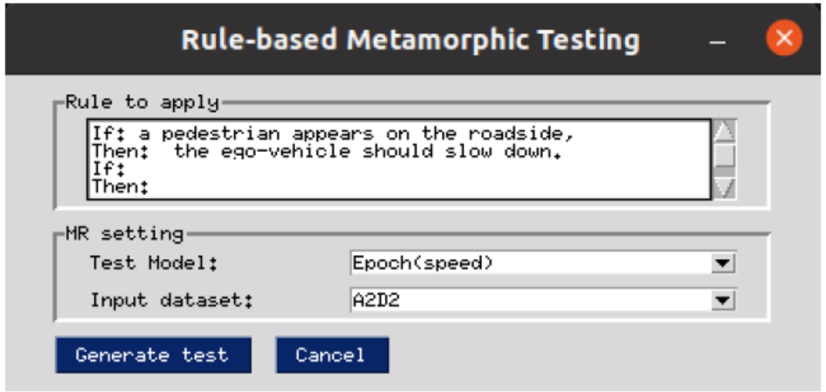
- **Image manipulation**
  - Implementing Transformation Add by masks extracted based on semantic segmentation

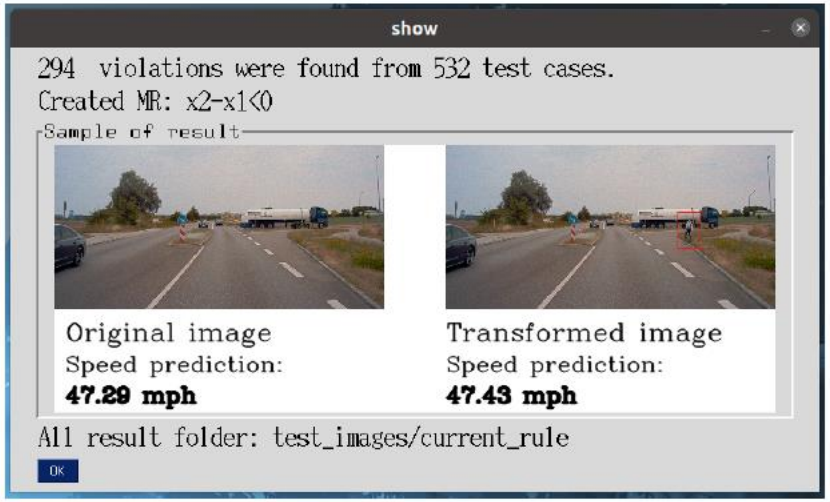
- **Image Synthesis**
  - Implementing Transformation Remove and Replace objects by Pix2pixHD GAN

- **Image-to-Image Translation**
  - Implementing Transformation Replace weather by UNIT GAN

Methodology

PROTOTYPE





GUI of the prototype

(a) Original driving scene (b) Original driving scene (c) Original driving scene (d) Original driving scene (e) Original driving scene

(f) Add a pedestrian (g) Add a slow sign (h) Remove a lane line (i) Replace buildings with trees (j) Transform to a night scene

Generated driving images

Results

RQ1: THE CAPABILITY OF RMT FOR VIOLATION DETECTION

	Epoch	VGG16	Resnet101
Rule 1	294 (55.26%)	46 (8.65%)	68 (12.78%)
Rule 2	244 (57.41%)	39 (9.18%)	63 (14.82%)
Rule 3	521 (97.93%)	509 (95.68%)	510 (95.86%)
Rule 4	239 (71.34%)	115 (34.33%)	74 (22.09%)
Rule 5	0 (0%)	76 (15.97%)	0 (0%)
Rule 6	9 (1.49%)	256 (42.38%)	0 (0%)
Rule 7	877 (93.10%)	273 (28.98%)	226 (23.99%)

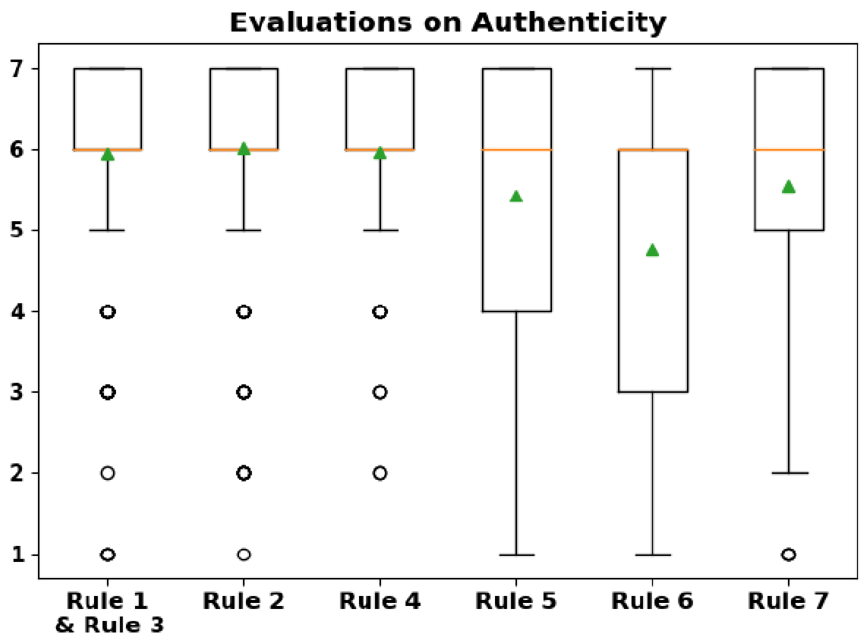
violations (ratio %) detected by seven rules

Rule	**Model**	**Threshold**					
		0	**10%**	**20%**	**30%**	**40%**	**50%**
Rule 1	**Epoch**	55.26%	96.80%	97.74%	97.93%	98.50%	98.68%
	VGG16	86.47%	88.72%	93.42%	95.68%	96.80%	97.18%
	Resnet101	12.78%	88.91%	94.92%	95.86%	96.62%	96.80%
Rule 2	**Epoch**	57.41%	97.65%	99.06%	99.29%	99.53%	99.76%
	VGG16	9.18%	89.18%	94.12%	96.47%	98.11%	98.59%
	Resnet101	14.82%	94.82%	96.47%	96.94%	97.18%	97.65%
Rule 5	**Epoch**	0	0	0	0	0	0
	VGG16	51.89%	15.97%	4.83%	1.68%	0.42%	0.21%
	Resnet101	0.21%	0	0	0	0	0
Rule 6	**Epoch**	4.30%	1.49%	0.50%	0	0	0
	VGG16	67.88%	42.38%	25.83%	14.90%	9.11%	5.46%
	Resnet101	2.81%	0	0	0	0	0
Rule 7	**Epoch**	93.10%	99.15%	99.36%	99.36%	99.47%	99.58%
	VGG16	28.98%	68.79%	86.84%	92.46%	94.28%	95.86%
	Resnet101	23.99%	26.75%	43.52%	62.42%	73.14%	78.56%

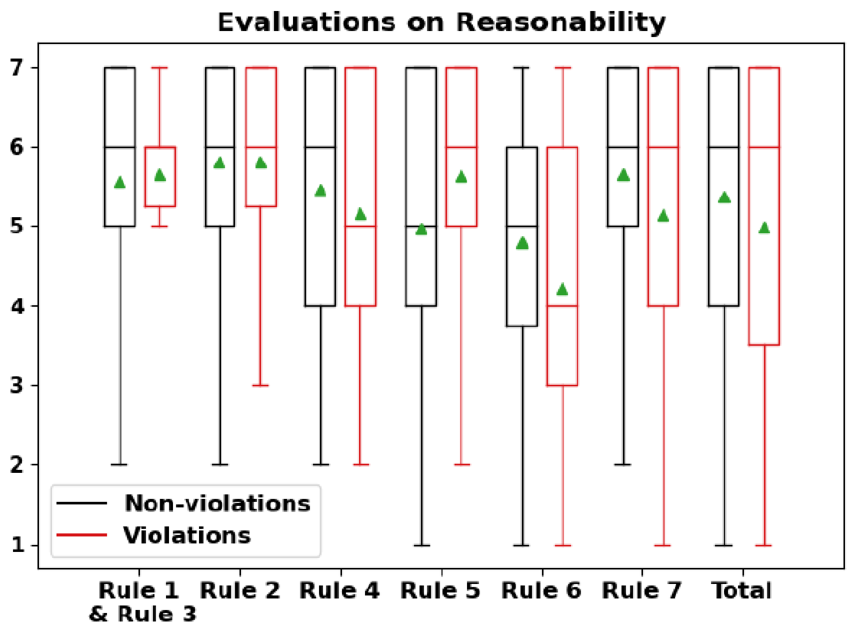
violations (ratio %) detected by rules with different thresholds

Results

RQ2 & RQ3: AUTHENTICITY OF GENERATED IMAGES AND VALIDITY OF DETECTED VIOLATION



Human assessment of the image authenticity in a 7-Point Likert Scale



Human assessment of the reasonability of new prediction in a 7-Point Likert Scale.

Results

RQ5: EFFICIENY OF RMT

	Image generation cost (s)	Evaluation cost (s)	Total cost (s)
Rules 1 & 3	137.50	5.58	143.08
Rule 2	131.30	4.80	137.10
Rule 4	279.17	4.80	283.97
Rule 5	196.67	4.78	201.45
Rule 6	195.07	6.08	201.15
Rule 7	126.74	4.37	131.11

Efficiency of RMT for rules

Results (Projects & Industry Perspective)

"Context-aware verification and validation framework for autonomous driving", CI, ARC Discovery Project grant DP210102447. Awarded $448,958. 2021-2024.

"A safety-preserving ecosystem for autonomous driving", Leader CI, ARC Linkage Project grant LP190100676. Awarded $341,853. 2021-2024.

NIO

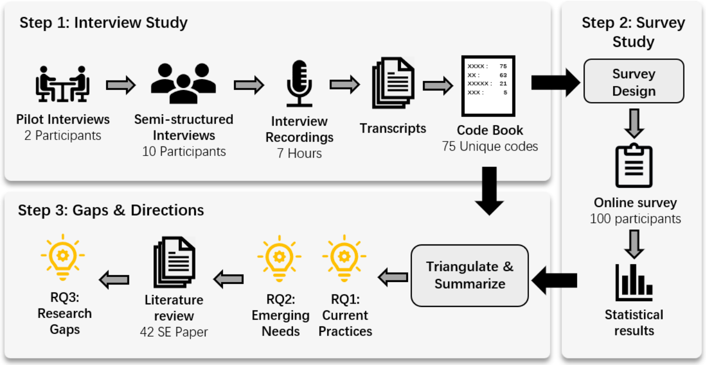
apollo

$$
\begin{aligned}
\mathit{scene} &:= \epsilon \mid \mathit{object} ; \mathit{scene} \\
\mathit{object} &:= \mathit{dynamic\ object} \mid \mathit{static\ object} \\
\mathit{dynamic\ object} &:= \langle \mathit{actor}, \mathit{action} \rangle \\
\mathit{actor} &:= \mathit{vehicle} \mid \text{pedestrian} \mid \mathit{cyclist} \mid \text{unknown} \\
\mathit{vehicle} &:= \text{truck} \mid \text{car} \mid \text{bus} \mid \text{van} \\
\mathit{cyclist} &:= \text{bicyclist} \mid \text{motorcyclist} \mid \text{tricyclist} \\
\mathit{action} &:= \text{stop} \mid \text{cruise} \mid \text{change lane} \mid \text{overtake} \mid \text{cross} \mid \ldots \\
\mathit{static\ object} &:= \mathit{traffic\ light} \mid \text{stop sign} \mid \text{crosswalk} \mid \text{intersection} \\
&\quad\ \text{traffic cone} \mid \ldots \mid \text{unknown} \\
\mathit{traffic\ light} &:= \langle \mathit{color}, \mathit{shape}, \mathit{orientation} \rangle \\
\mathit{color} &:= \text{red} \mid \text{green} \mid \text{yellow} \mid \text{black} \\
\mathit{shape} &:= \text{square} \mid \text{round} \\
\mathit{orientation} &:= \text{vertical} \mid \text{horizontal}
\end{aligned}
$$

https://apolloscape.auto/

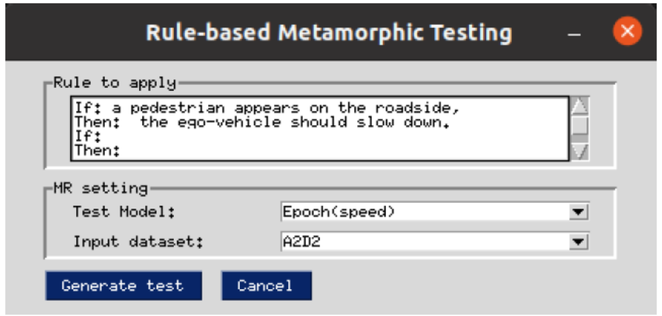
Talk Agenda

A Study



Test Reduction

Test Generation



Future/Ongoing Work

An Analysis of Adversarial Attacks and Defenses on Autonomous Driving Models (PerCom’20)

- Attacks (White Box and Black Box)

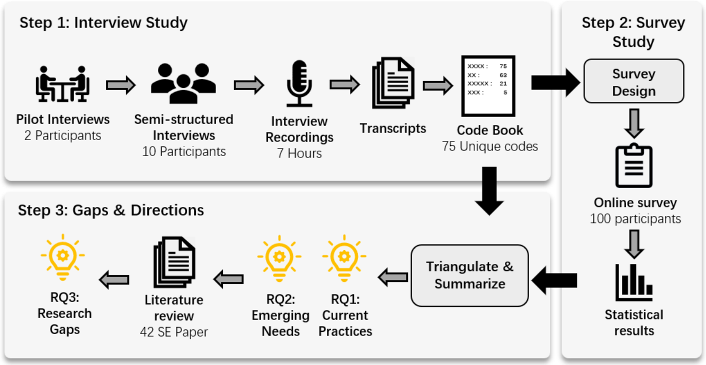
- Defences
  - Defensive distillation
  - Adversarial training
  - Feature squeezing
  - Anomaly Detection

[1] Y Deng, X Zheng, TY Zhang, C Chen, GN Lou, M Kim. "An Analysis of Adversarial Attacks and Defenses on Autonomous Driving Models" IEEE International Conferene on Pervasive Computing and Communications (PerCom)’20

MSF Robustness

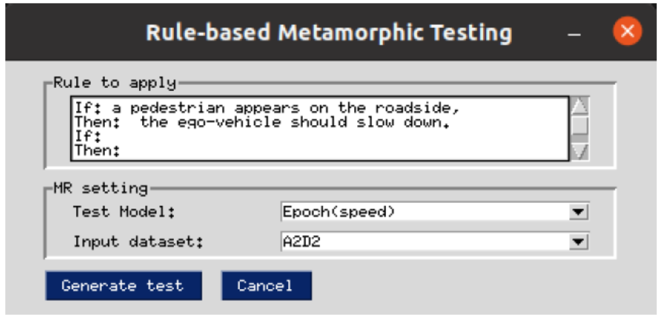
Talk Agenda

A Study



Test Reduction

Test Generation



Future/Ongoing Work

UAV Auto Landing System – Marker: Robustness issues

(a) Ground (b) Lawn

Fig. 6. Example of test map in real-world (First row) and AirSim (Second row).

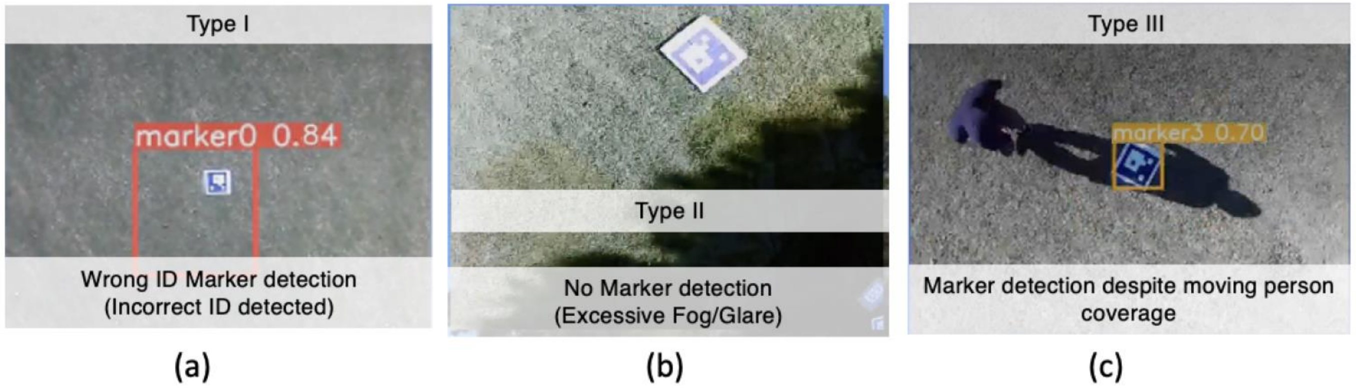


Fig. 11. Example of our real-world violation reproduction, (a) a false positive detection case, (b) failure to land because no object is detected as a marker, and (c) is a violation due to obstruction that highlights the detected marker after perturbation leading to a potential collision.

Liang, L., Deng, Y., Morton, K., Kallinen, V., James, A., Seth, A., Kuantama, E., Mukhopadhyay, S., Han, R. and Zheng, X., 2023. RLaGA: A Reinforcement Learning Augmented Genetic Algorithm For Searching Real and Diverse Marker-Based Landing Violations. *arXiv preprint arXiv:2310.07378*.

UAV Auto Landing System: Markerless and Moving objects (Challenges in testing and verification)

https://yachtharbour.com/news/what-you-need-to-know-about-superyacht-helipads-3229?src=news_view_page_bar

"Robust and Scalable Autonomous Landing for Drones", CI, ARC Linkage Project grant LP210100337. Awarded $459,593. 2022-2025.

MACQUARIE University

Formal Guarantee for Learning-enabled CPS

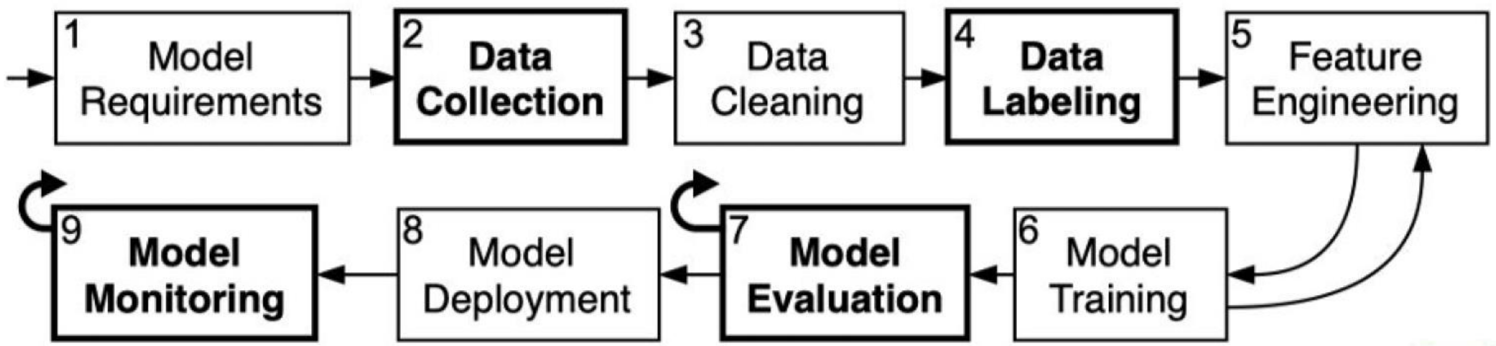


Figure 1: Parallel workflow for ML [3]

https://www.tacps.org/

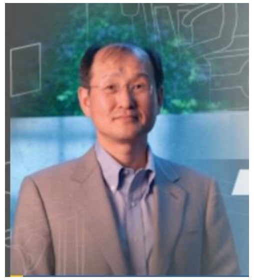
Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. Software engineering for machine learning: A case study. In 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP). IEEE, 291–300.

Organizing Committee members and Selected Keynote Speakers - TACPS

Xi Zheng

MACQUARIE University

Aloysius Mok

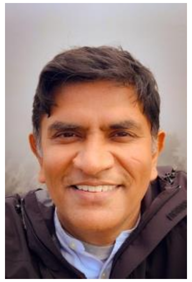


Insup Lee

Penn
University of Pennsylvania

Oleg Sokolsky

Penn
University of Pennsylvania



Bhaskar Krishnamachan

USC

Dakai Zhu

Ruzica Piskac

Yale

Sanjit Seshia

Berkeley
UNIVERSITY OF CALIFORNIA

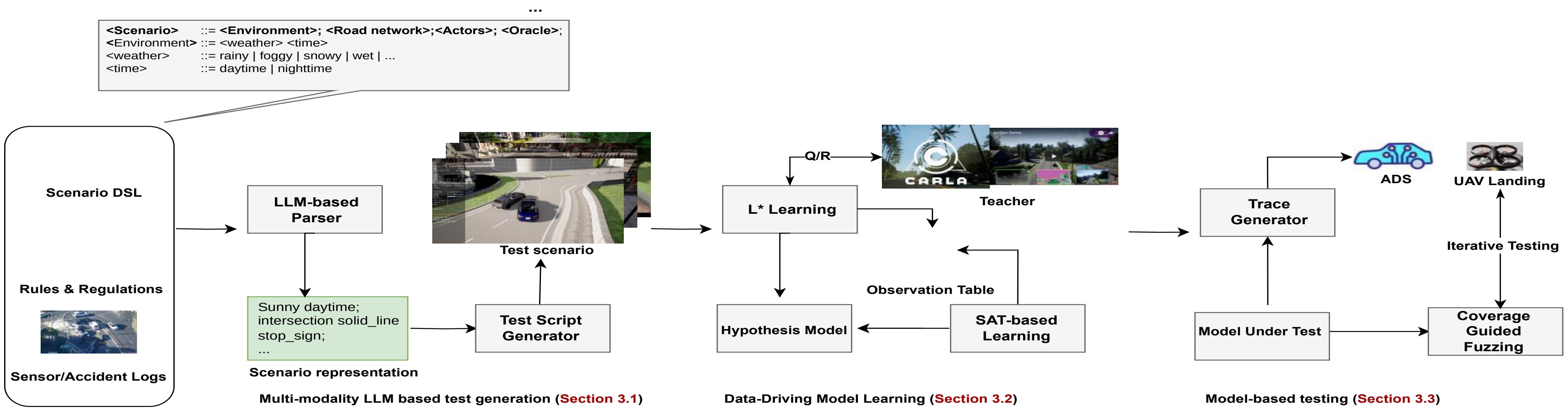
Tevfik Bultan

Paulo Tabuada

Ufuc Topcu

Part of the white-paper steering committee members – Inaugural TACPS in San Diego (25-26th Jan 2024)

Ongoing Work – Towards Formal Testing (Vision Paper)



Zheng, X., Mok, A.K., Piskac, R., Lee, Y.J., Krishnamachari, B., Zhu, D., Sokolsky, O. and Lee, I., 2024, July. Testing learning-enabled cyber-physical systems with large-language models: A formal approach. In Companion Proceedings of the 32nd ACM International Conference on the Foundations of Software Engineering (pp. 467-471).

Penn UNIVERSITY of PENNSYLVANIA

Yale University

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

TARGET: Traffic Rule-based Test Generation for Autonomous Driving Systems – LLM knowledge extraction pipeline

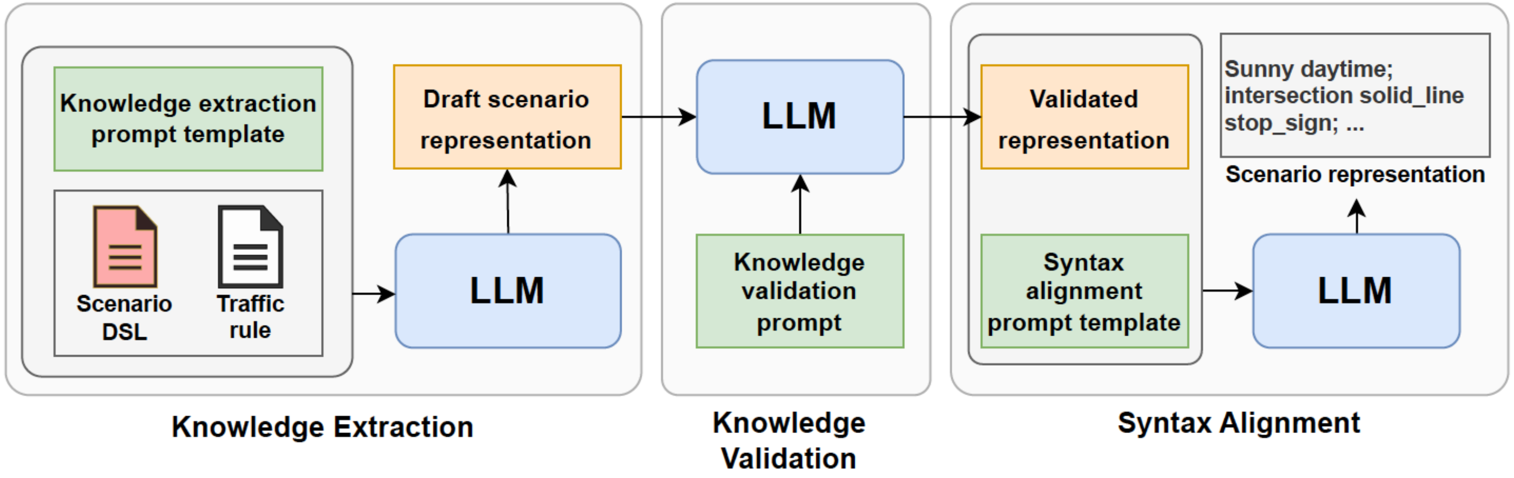


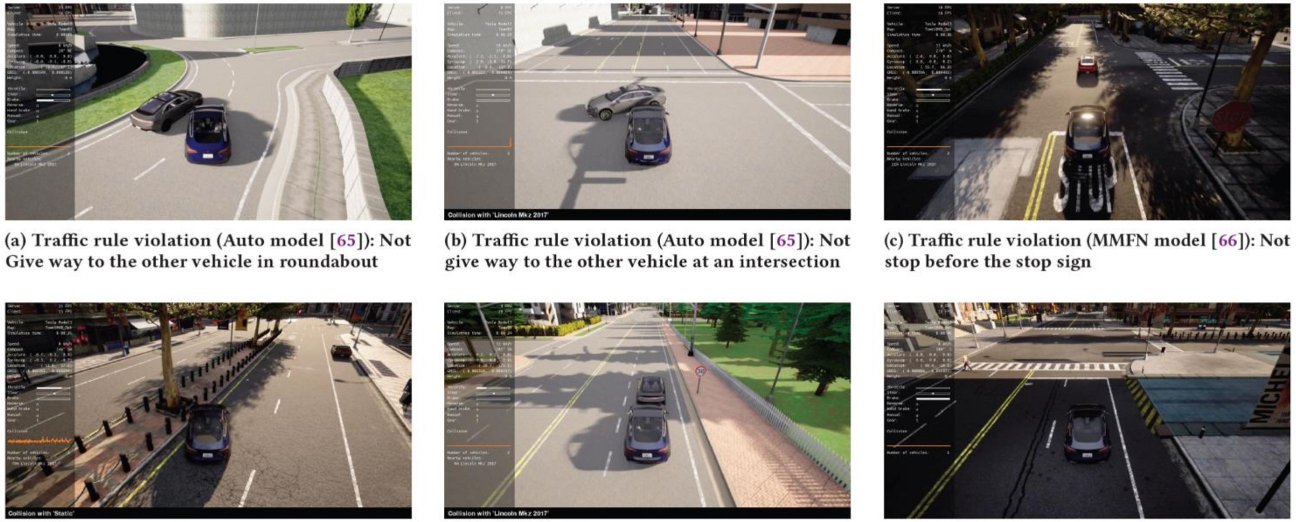
Fig. 3. The workflow of traffic rule parsing

Deng, Y., Yao, J., Tu, Z., Zheng, X., Zhang, M. and Zhang, T., 2023. TARGET: Traffic Rule-based Test Generation for Autonomous Driving Systems. arXiv preprint arXiv:2305.06018.

Target – Test Generation using LLM

Role Setting
You are a test expert for autonomous driving systems. Your task is to generate specific test scenario representations from given traffic rules.
Prompt
Below is the definition of a domain-specific language to represent test scenarios for autonomous driving systems: {*Details of DSL*} Below are the lists of commonly used elements for each subcomponent. When creating the scenario representation, consider the following elements first for each subcomponent. If no element can describe the close meaning, create a new element by yourself. {*Details of element lists*} Below is an example of an input traffic rule and the corresponding scenario representation: Traffic rule: {*content of the traffic rule*} Scenario representation: {*content of the scenario representation*} Based on the above descriptions and examples, convert the following traffic rule to corresponding scenario representation: {*input traffic rule*}

Table 1: The prompt template for knowledge extraction



(a) Traffic rule violation (Auto model [65]): Not Give way to the other vehicle in roundabout
(b) Traffic rule violation (Auto model [65]): Not give way to the other vehicle at an intersection
(c) Traffic rule violation (MMFN model [66]): Not stop before the stop sign
(d) MMFN model [66] collides on the roadside
(e) Autoware [28] collides with the front vehicle
(f) LAV [11] does not move even if the pedestrian has crossed the road

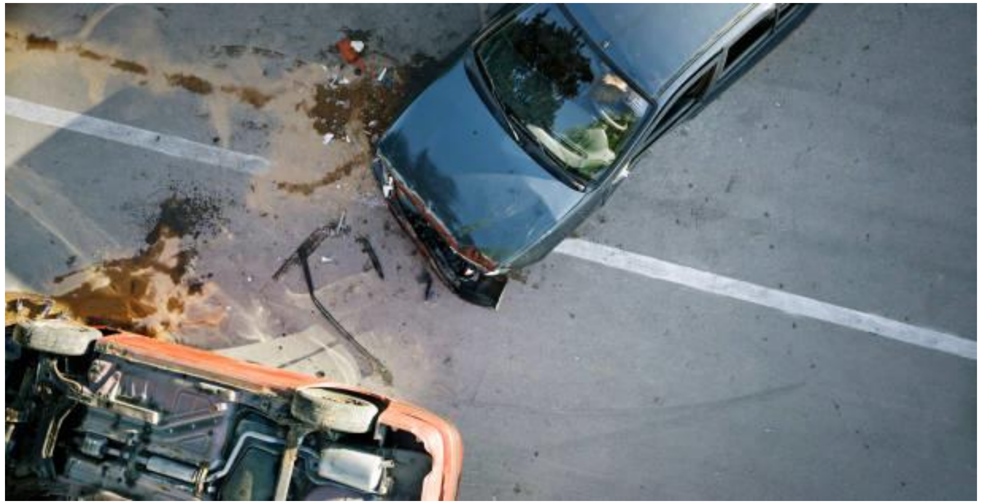
Figure 2: Examples of detected problems on ADSs and Carla simulator

PURDUE UNIVERSITY

Meta

Deng, Y., Yao, J., Tu, Z., Zheng, X., Zhang, M. and Zhang, T., 2023. TARGET: Traffic Rule-based Test Generation for Autonomous Driving Systems. *arXiv preprint arXiv:2305.06018*.

Ongoing Work – Accidents Analysis using Multi-modal LLM



Q: Can you imagine what leads to such a collision?

A: There are several factors that could lead to a collision like the one depicted in the image. Some common causes include: 1) Speeding; 2) Distracted driving;
3) Failure to yield the right of way; 4) Poor visibility; 5) Mechanical issues; and 6) Road conditions.

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Future Work – Industry adaptation of model-based testing & Neuro-symbolic paradigm

apollo

KONVERY
恺望数据

Penn
UNIVERSITY of PENNSYLVANIA

LUX ET VERITAS
YALE
Yale University

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

THE UNIVERSITY OF CALIFORNIA
LET THERE BE LIGHT
UCLA

Future Work – Industry Deployment

SAVION
AEROSPACE

Collaborators (with publications)

Christine Julien

THE UNIVERSITY OF TEXAS AT AUSTIN

Miryung Kim

THE UNIVERSITY OF CALIFORNIA UCLA

Tianyi Zhang

PURDUE UNIVERSITY

Mengshi Zhang

TY Chen

SWINBURNE UNIVERSITY OF TECHNOLOGY

Michael Sheng

MACQUARIE University

Yun Yang

SWINBURNE UNIVERSITY OF TECHNOLOGY

Richard Han

MACQUARIE University

Wanlei Zhou

澳門城市大學 City University of Macau

Guangquan Xu

TIANJIN UNIVERSITY (PEIYANG UNIVERSITY) 1895

Tian Wang

BEIJING NORMAL UNIVERSITY 1902

Jun Wu

SHANGHAI JIAO TONG UNIVERSITY 1896

Shui Yu

UTS

Yang Xiang

SWINBURNE UNIVERSITY OF TECHNOLOGY

Chen Wang

DATA 61

Xiao Liu

DEAKIN UNIVERSITY

Huai Liu

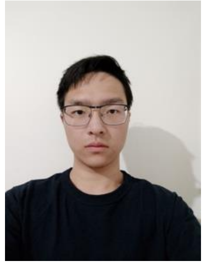
SWINBURNE UNIVERSITY OF TECHNOLOGY

Yipeng Zhou

MACQUARIE University

Mentees

Autonomous systems Testing and Verification



Yao Deng

MACQUARIE University

Jiaohong Yao

MACQUARIE University

Siwei Luo

MACQUARIE University

Linfeng Liang

MACQUARIE University

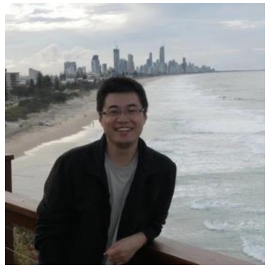
Guannan Lou

MACQUARIE University

MSF Robustness Analysis

Chuxuan Tong

MACQUARIE University



Jiwei Guan

MACQUARIE University

Distributed Learning and knowledge-guided ML training

Sangsu Lee

Steven Yu

Jianchao Lu

MACQUARIE University

Yupeng Jiang

MACQUARIE University

Yi He

MACQUARIE University

Jiaqi Ge

MACQUARIE University

Yong Li

MACQUARIE University

Questions ?