Commonwealth Cyber Initiative

Towards Trustworthiness in Autonomous Vehicles

Evgenia Smirni (PI)
William and Mary
Homa Alemzadeh (coPI)
UVA

Xugui Zhou, Anna Schmedding, Haotian Ren, Lishan Yang, Yiyang Lu (graduate students)
P. Schowitz (undergraduate student)

WILLIAM & MARY
CHARTERED 1693

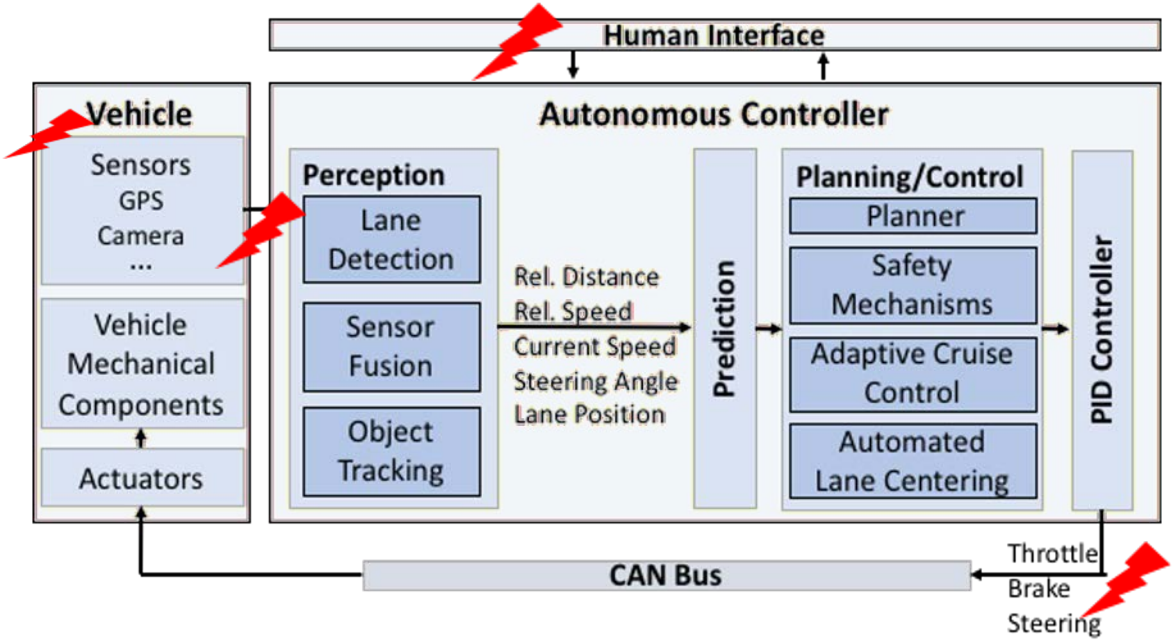
UNIVERSITY of VIRGINIA

We rely on advanced driver assistance systems (ADAS) to improve every-day driving experience but... are they safe?

Commonwealth Cyber Initiative

Simulation: Carla w/ OpenPilot

- Carla driving simulator
- ADAS: OpenPilot
  - level-2 autonomy: automated lane centering, adaptive cruise control, lane change assistance
- Complicated/vulnerabilities
- Manipulate control commands
- **Context-aware attack**
  - strategic!
  - gas/break/steering



Commonwealth Cyber Initiative

Impact: Car Safety

	Alerts	Hazards	Accidents	Hazards & No Alerts	Time to Hazard
No Attacks	0.1%	0.0%	0.0%	0.0%	N/A
Context-Aware Attack	9.7%	96.3%	62.1%	86.7%	2.3 seconds

- Observations
  - Lane invasions are common
  - Context-aware attack: effective
    - start time/duration
  - Human alertness
  - Steering attack: most vulnerable

Paper appeared at DSN 2022, June 2022, acceptance rate 18%,

"Strategic Safety-Critical Attacks Against an Advanced Driver Assistance System"

Commonwealth Cyber Initiative

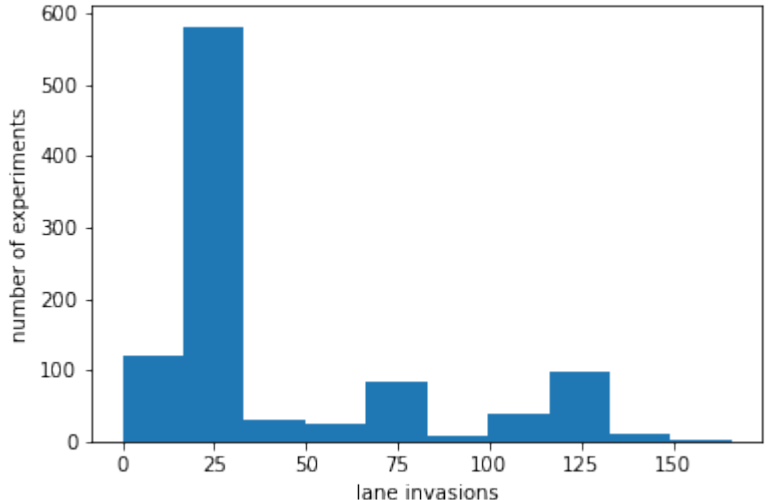
Closer Look: ML Components…

- Neural Networks for image recognition
- What happens if bit flips happen?
- Can a simple bit flip in a ML component of OpenPilot create hazards?
- Other self-driving software? LBC

Supercombo Layer 1 Fault Injection

- **Weather:** Cloudy
- **Number of experiments:** 1000

- Many lane invasions
- 7 Hazards flagged (e.g. collision with curb)



Commonwealth Cyber Initiative

Learning by Cheating: LBC

Inputs:

- Image
- Velocity
- Command vector (high level)
  - [Follow-lane, Turn-left, Turn-right, Go-straight

Outputs

- Waypoints (x5)

VERY EASY to cause collisions!

Open Problems…

- How resilient are NNs in safety-critical systems, specifically AVs?
  - Most faults have negligible effects
  - Some are critical
- Are certain parts of the NN more vulnerable than others? Are parts of the NN that tie in other inputs more vulnerable?
  - Early layers
- Where do bit-flips need to be to cause safety hazards?
  - Exponent bits have bigger impact
- Do certain times or actions performed by other parts of the system affect how vulnerable the NN is?
  - Turns
- Do environmental conditions affect how vulnerable the NN is?
  - Bad weather (e.g. rain) makes the NN more vulnerable
- Are certain contexts vulnerable regardless of the NN deployed?

CNNProtector: Protecting CNNs in Safety-Critical Systems

Protecting CNNs in Safety-Critical Systems

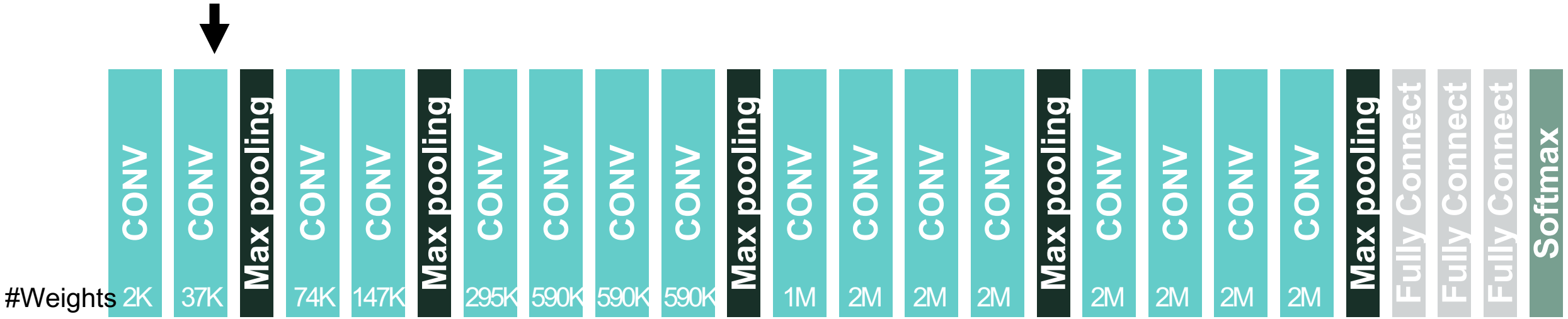
- Strict reliability constraints
- Reliability must be easy to incorporate
- Strict overhead constraints

Fault Model

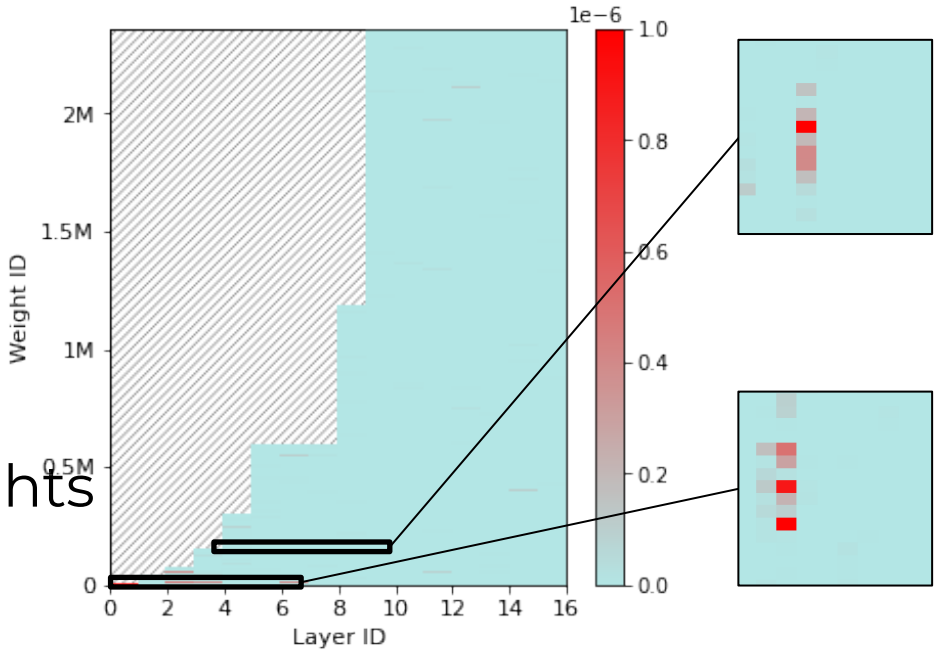
- Transient faults (soft errors) in DRAM
- Double bit-flip in one 32-bit floating point weight
- Cannot be corrected by ECC

Challenges: too many weights!

VGG19, Incepetion_v3, and ResNet50



Are some weights more important than others?



Importance scores

- Too many fault sites to search
- Pruning – assigns importance scores to weights
- Protect only the most important weights
  - *Taylor Pruning*

VGG19 observations

- Early layers more vulnerable
- For different datasets, different important weights

Summary

CNNs are vulnerable and need protection.

Our new tool, *CNNProtector* has:

- ✓ Low runtime overhead
- ✓ Low memory overhead
- ✓ Easy to use
  - Few lines of code for developers
  - No advanced knowledge of reliability needed
- ✓ Open source

Evgenia Smirni
William and Mary
esmirni@cs.wm.edu