Certified Perception for Autonomous Cars

MT-CPS · May 18, 2021

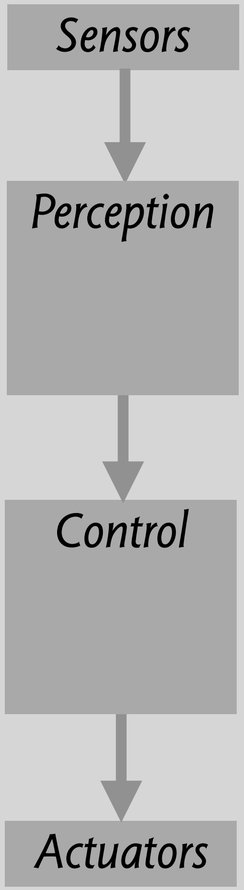
Uriel Guajardo, Annie Bryan, Nikos Arechiga,, Sergio Campos, Jeff Chow,, Daniel Jackson, Soonho Kong, Geoffrey Litt, **Josh Pollock**

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 1745302.

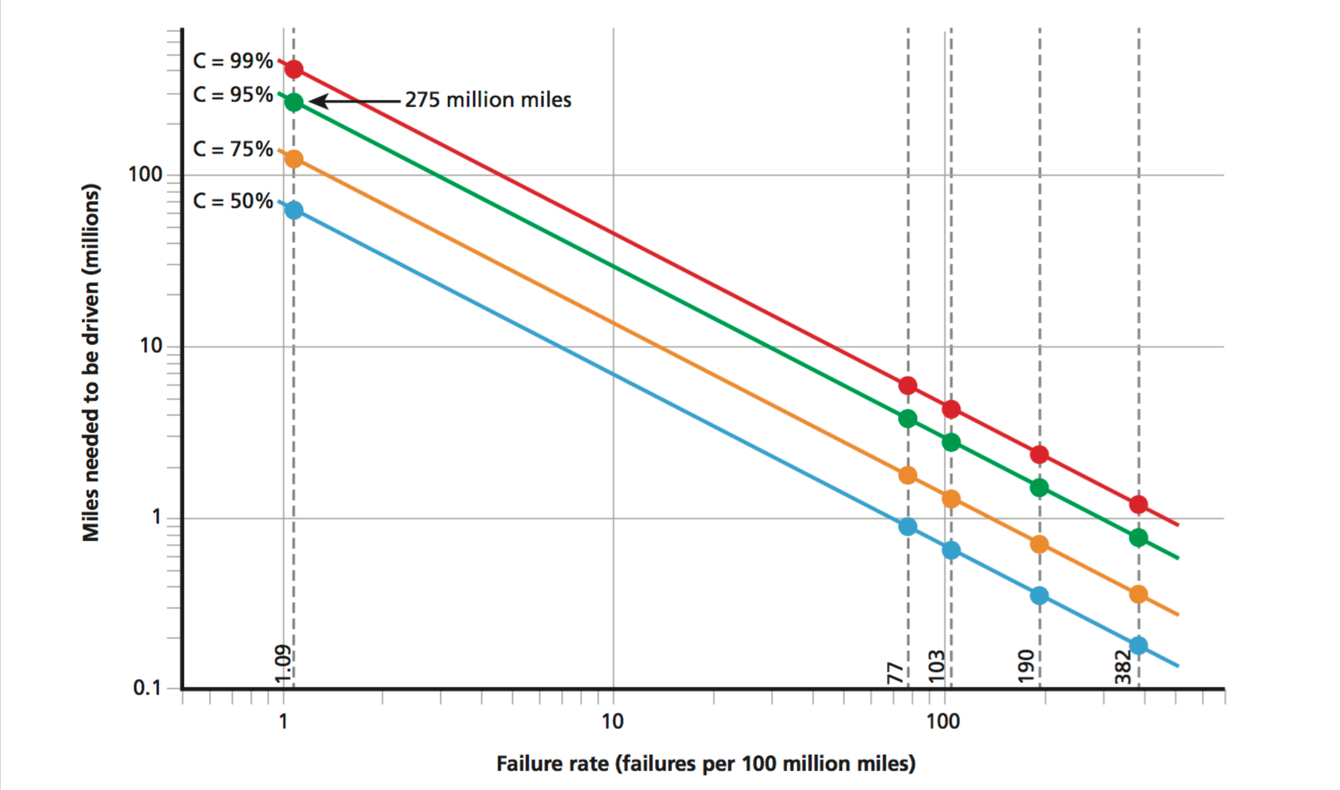
first pedestrian fatality (uber, tempe AZ, march 2018)

first pedestrian fatality (uber, tempe AZ, march 2018)

perception errors propagate through an autonomous driving system (ADS)



what about testing?



for 95% confidence in 1 death/100M miles,
need to drive 275M miles [RAND]

Waymo: 10M miles/year

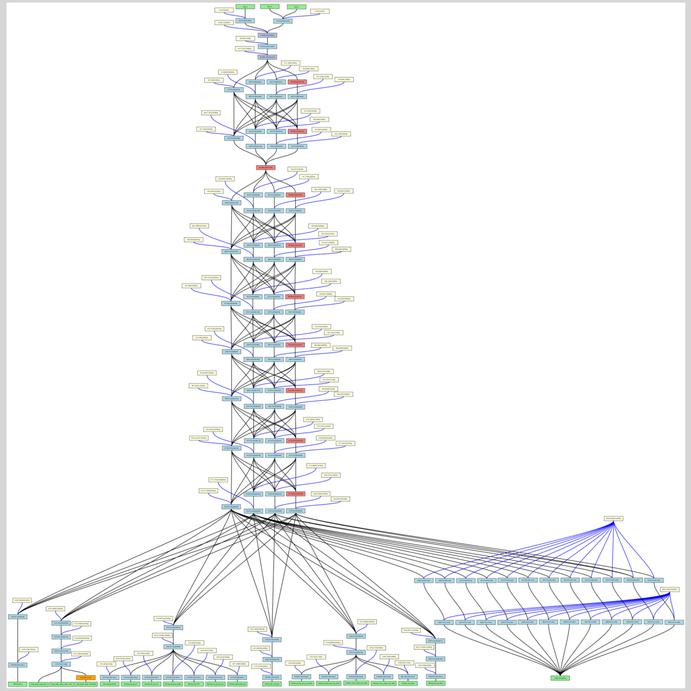
AND

must retest after every modification or upgrade

needs too much data

what about static verification?

SEL4 verification: $400/line
1m lines of critical code: $400M

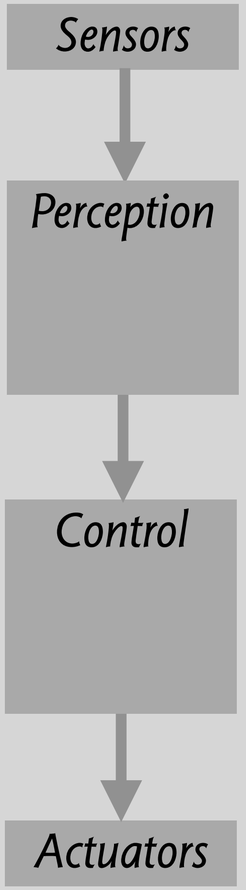


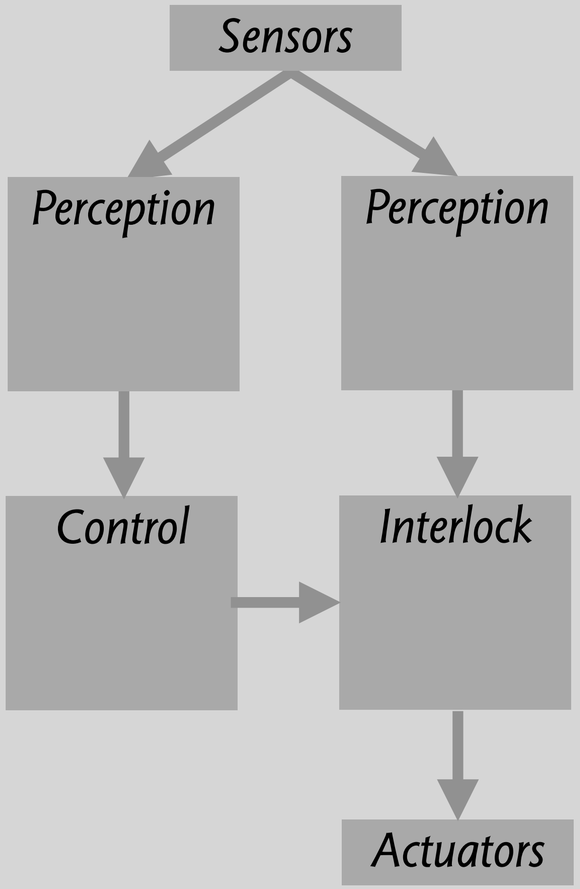
too expensive

too complex

https://teslamotorsclub.com/tmc/threads/tesla-autopilot-hw3.139550/

interlocks





no interlock

classic interlock

classic interlock leads to too many false positives

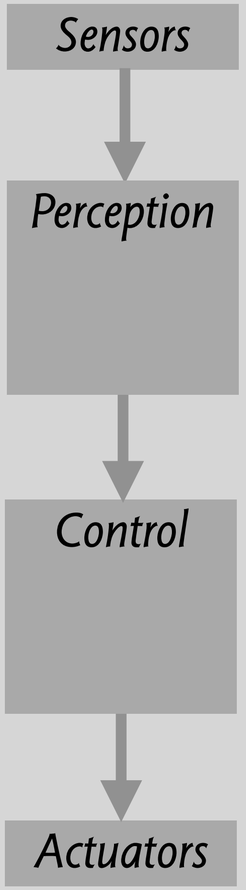
Uber Autopilot

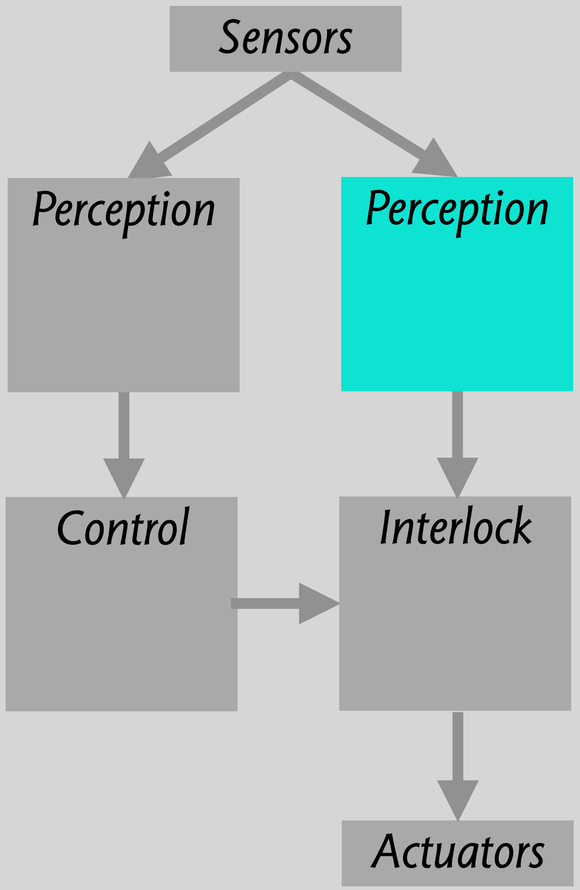
Volvo Driver-Assist

classic interlock leads to too many false positives

Uber Autopilot

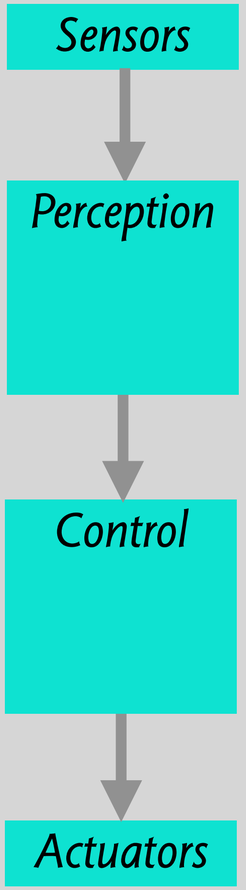
Volv… Dri… sist



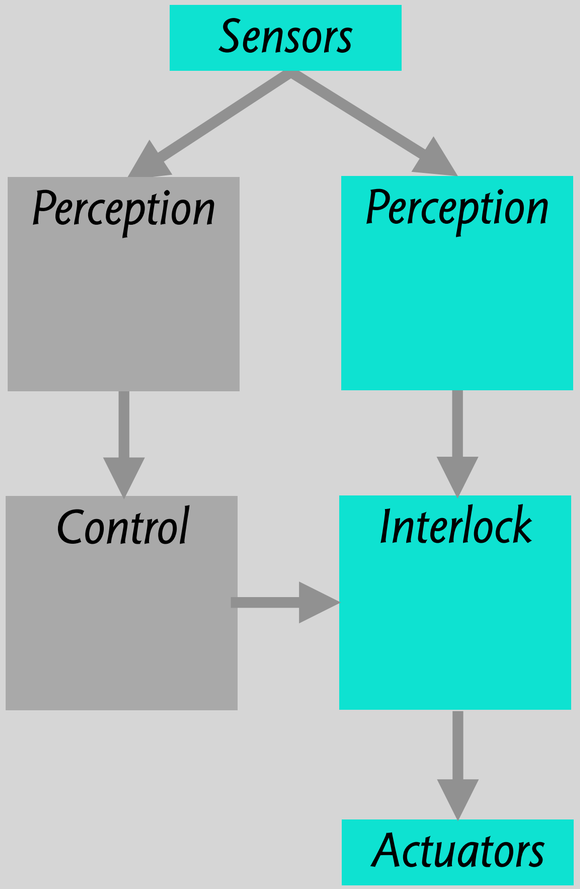


no interlock

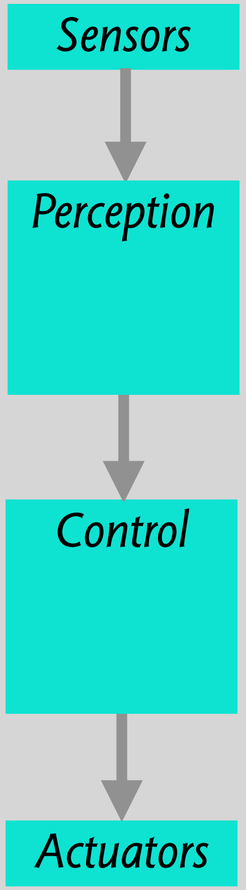
classic interlock



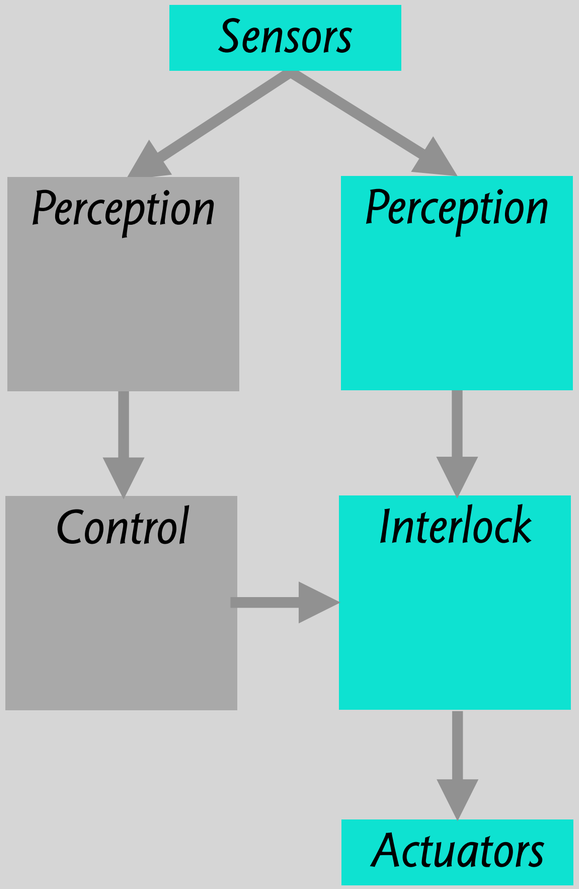
no interlock



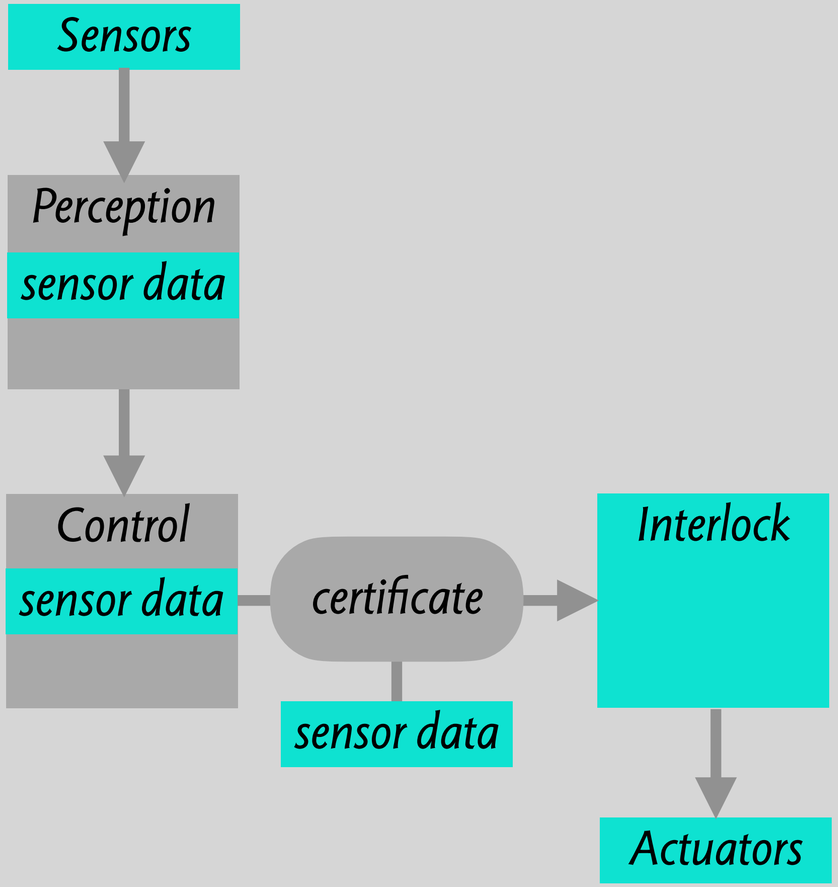
classic interlock



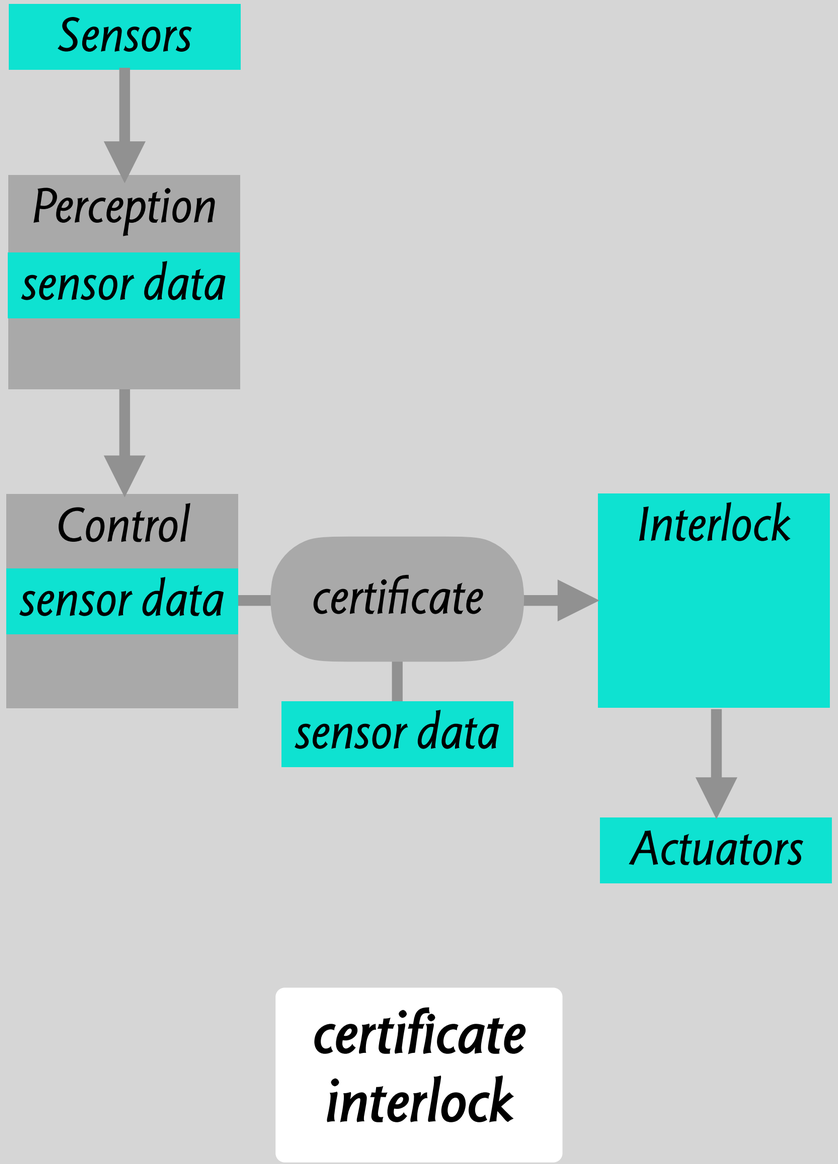
no interlock



classic interlock



certificate interlock



certificate interlock

- perception-control system *constructs* evidence
- system presents evidence *certificate* to interlock
- interlock *checks* certificate

trusted base

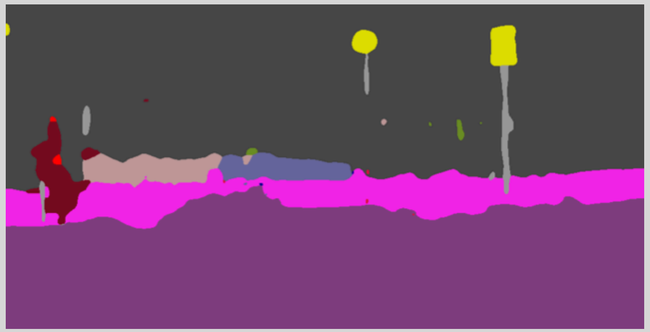
- sensors and actuators
- signed sensor data
- interlock

a segmentation certificate

adversarial segmentation example

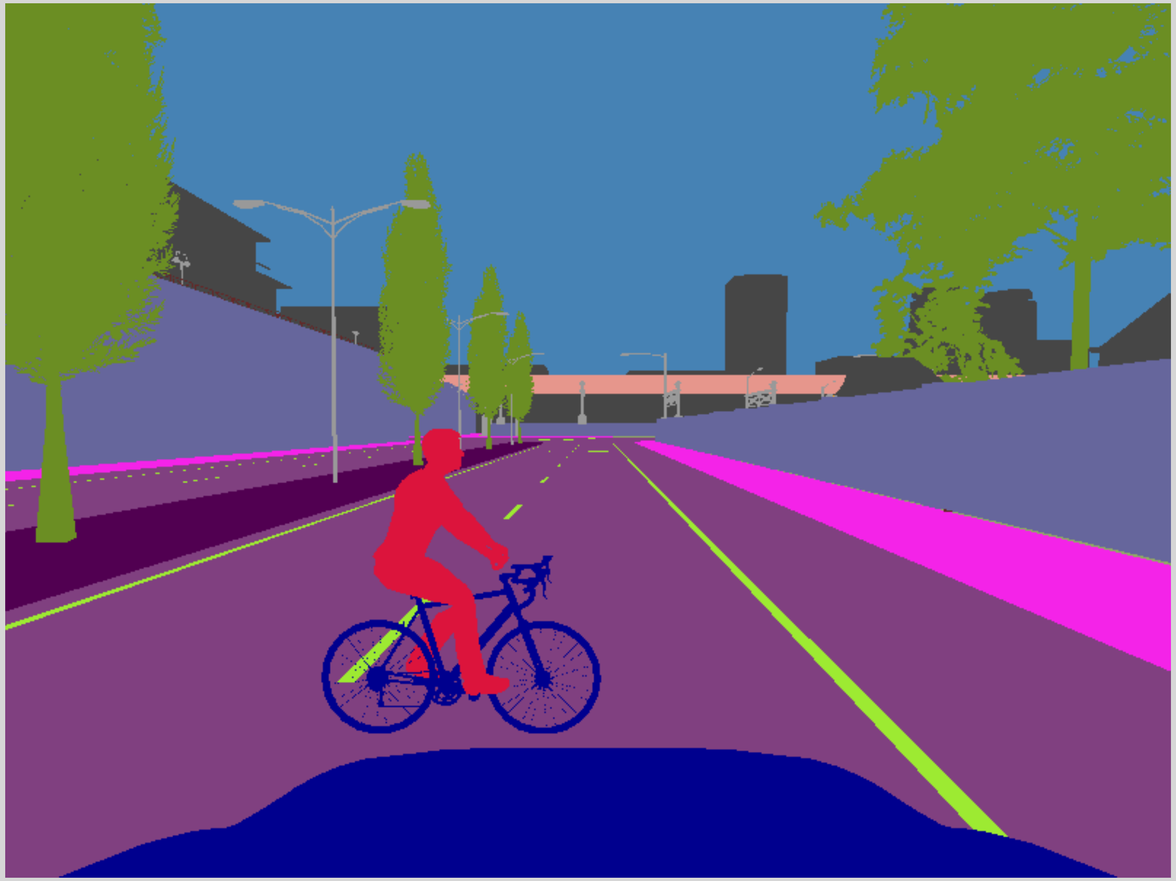
original image

adversarial image



Volker Fischer, Mummadi Chaithanya Kumar, Jan Hendrik Metzen & Thomas Brox.
Adversarial Examples For Semantic Image Segmentation (ICLR 2017)

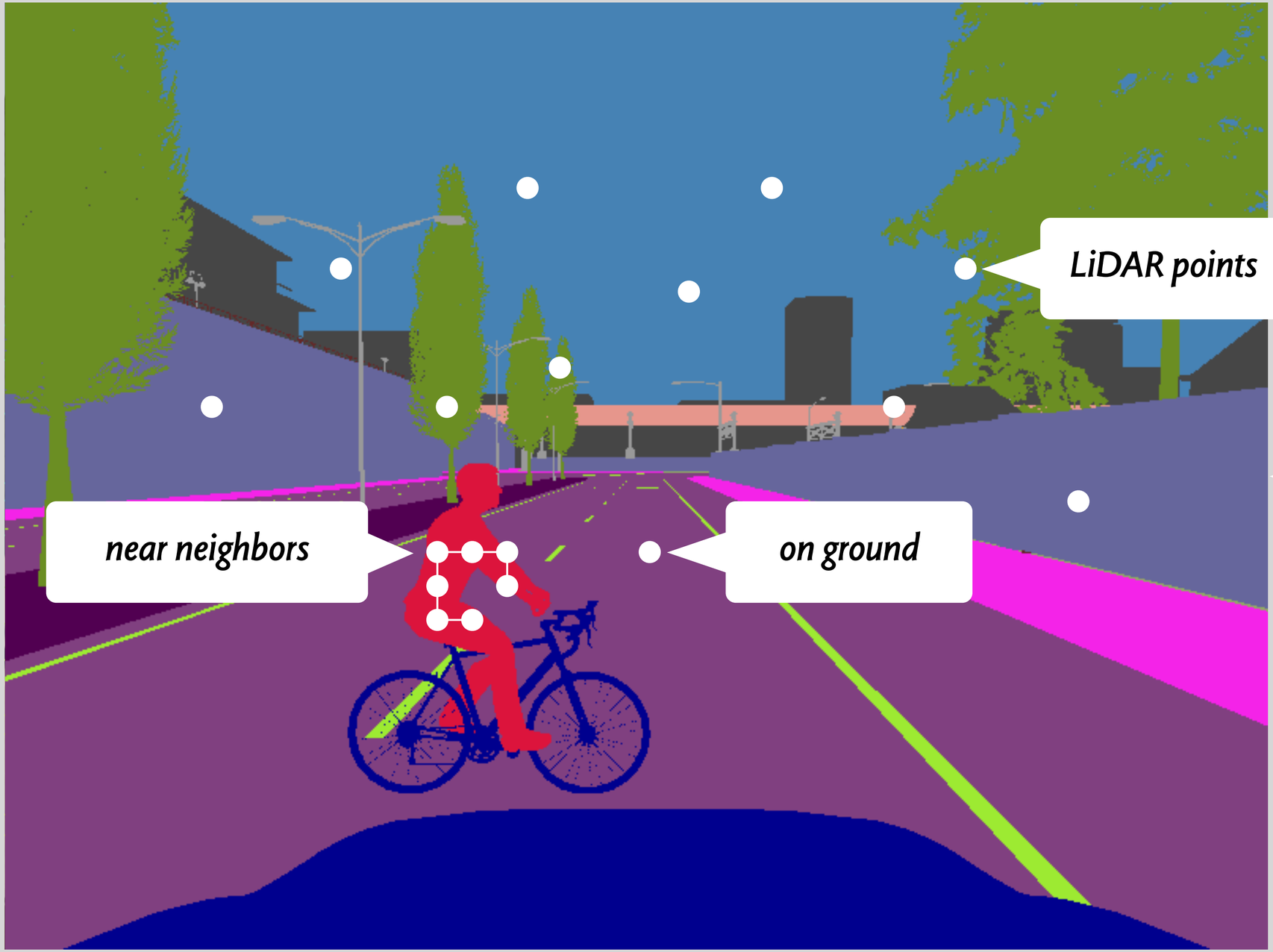
CARLA's segmentation



certificate
data

certificate data

certificate data



a segmentation certificate

certificate data

all points: a set of LiDAR points with position & velocity

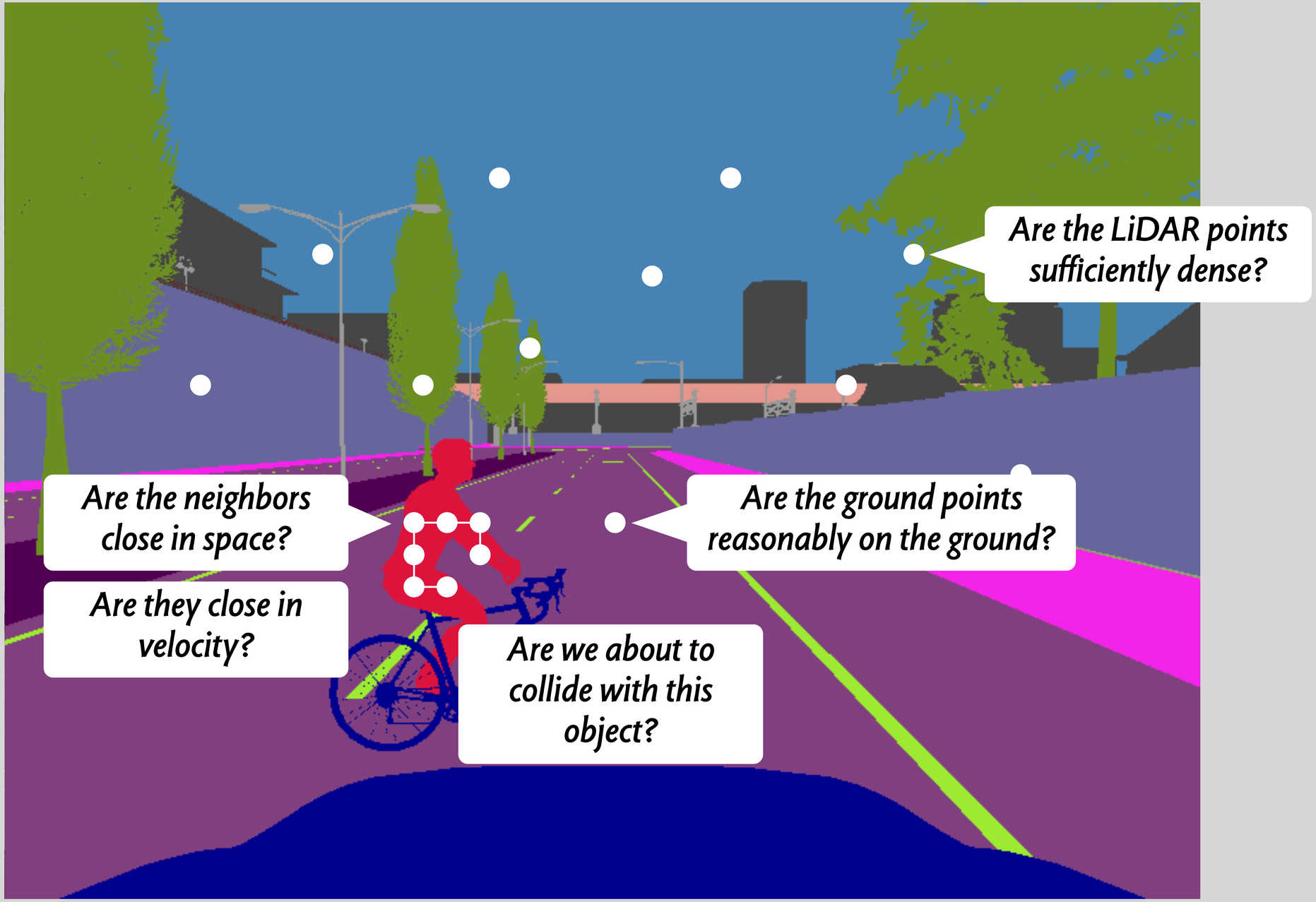
segmentation: a point -> object id map

ground ids: a list of object ids for ground plane objects

traversal: for each object id, a list of point pairs

certificate checks

certificate checks



a segmentation certificate

certificate checks

sufficient density: input LiDAR points are sufficiently dense

ground height: all ground points are close to the plane below the car

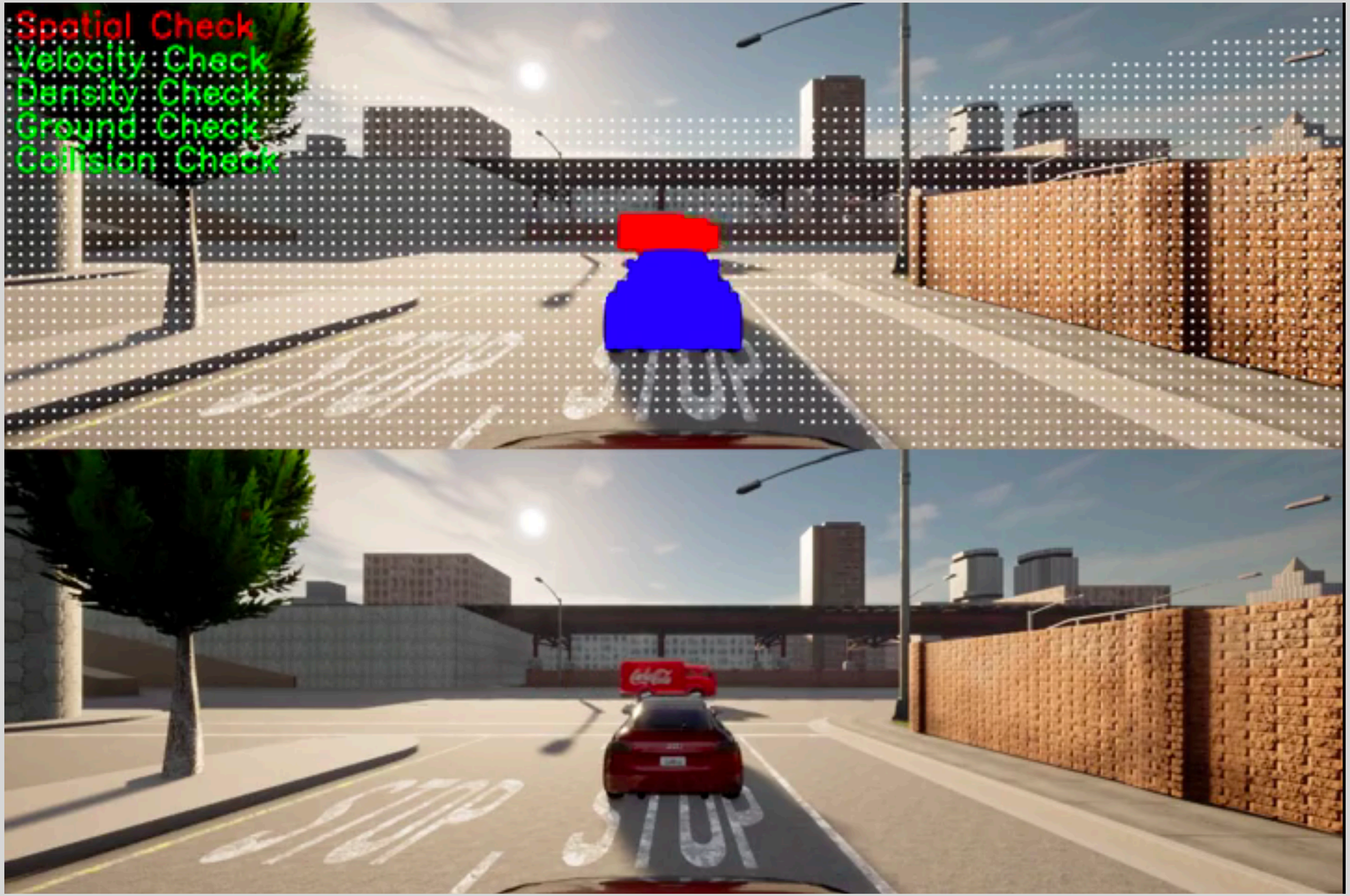
spatial contiguity: in each traversal, in each pair, the two points are close

consistent velocity: in each traversal, in each pair, the two points have similar velocities

collision prediction: the car will not collide with non-ground objects in the near future

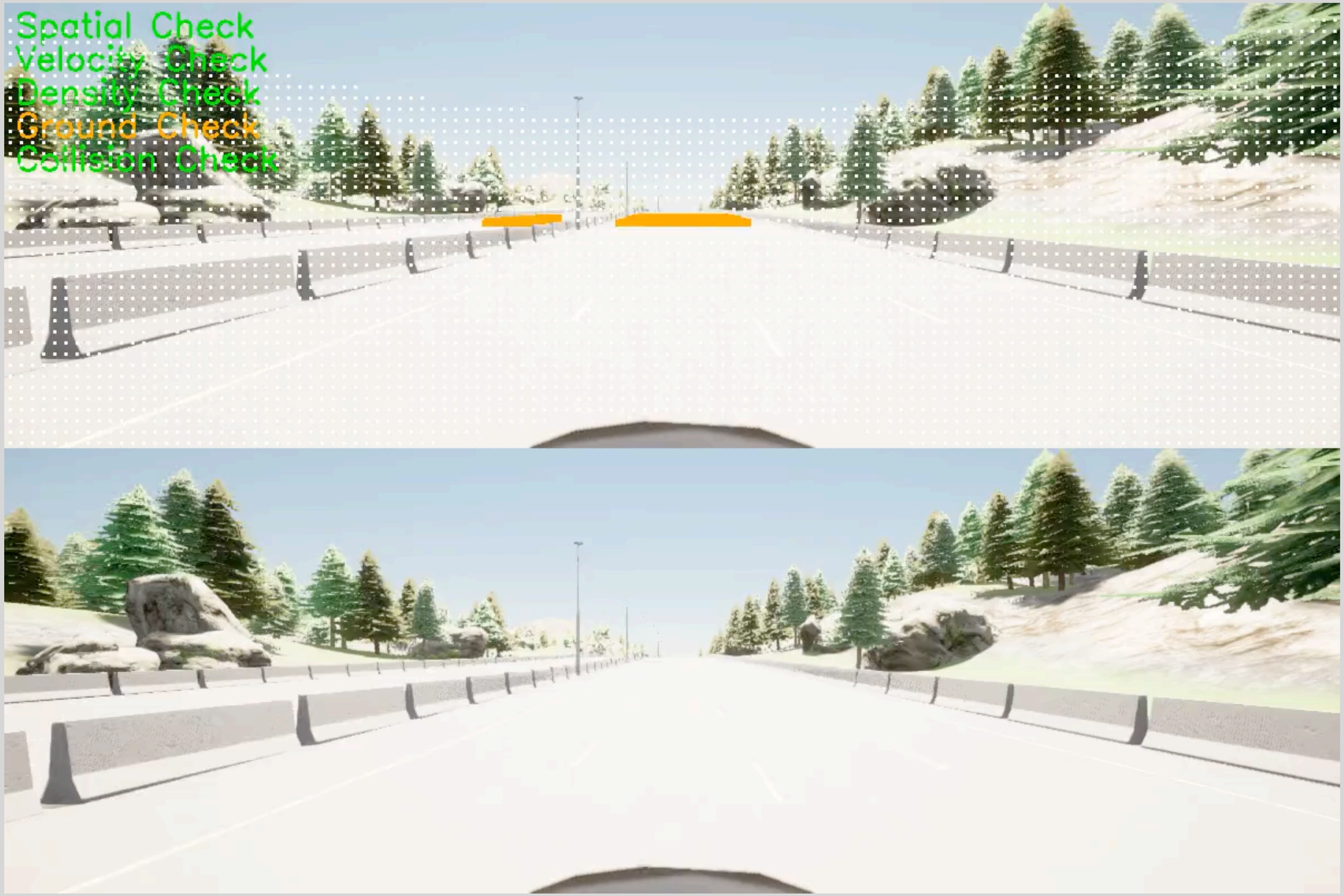
experiments

detecting a conflation of two objects

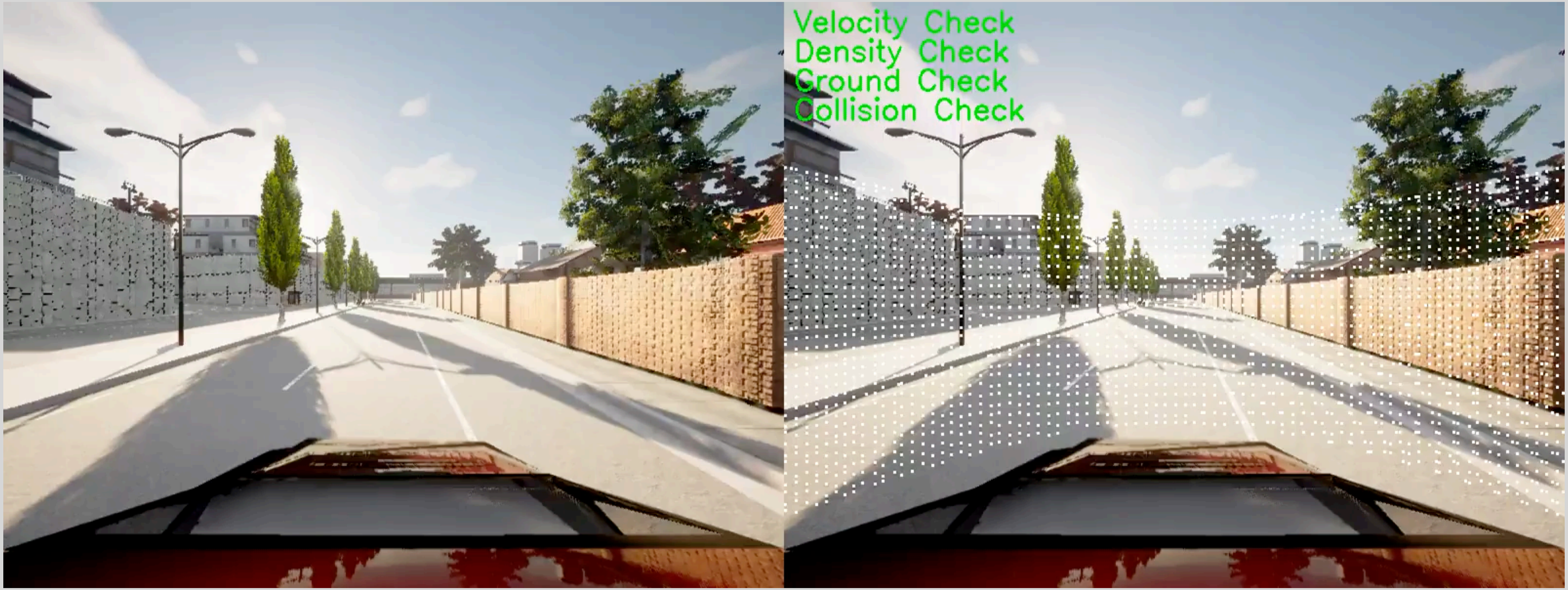


maintaining separation from car in front

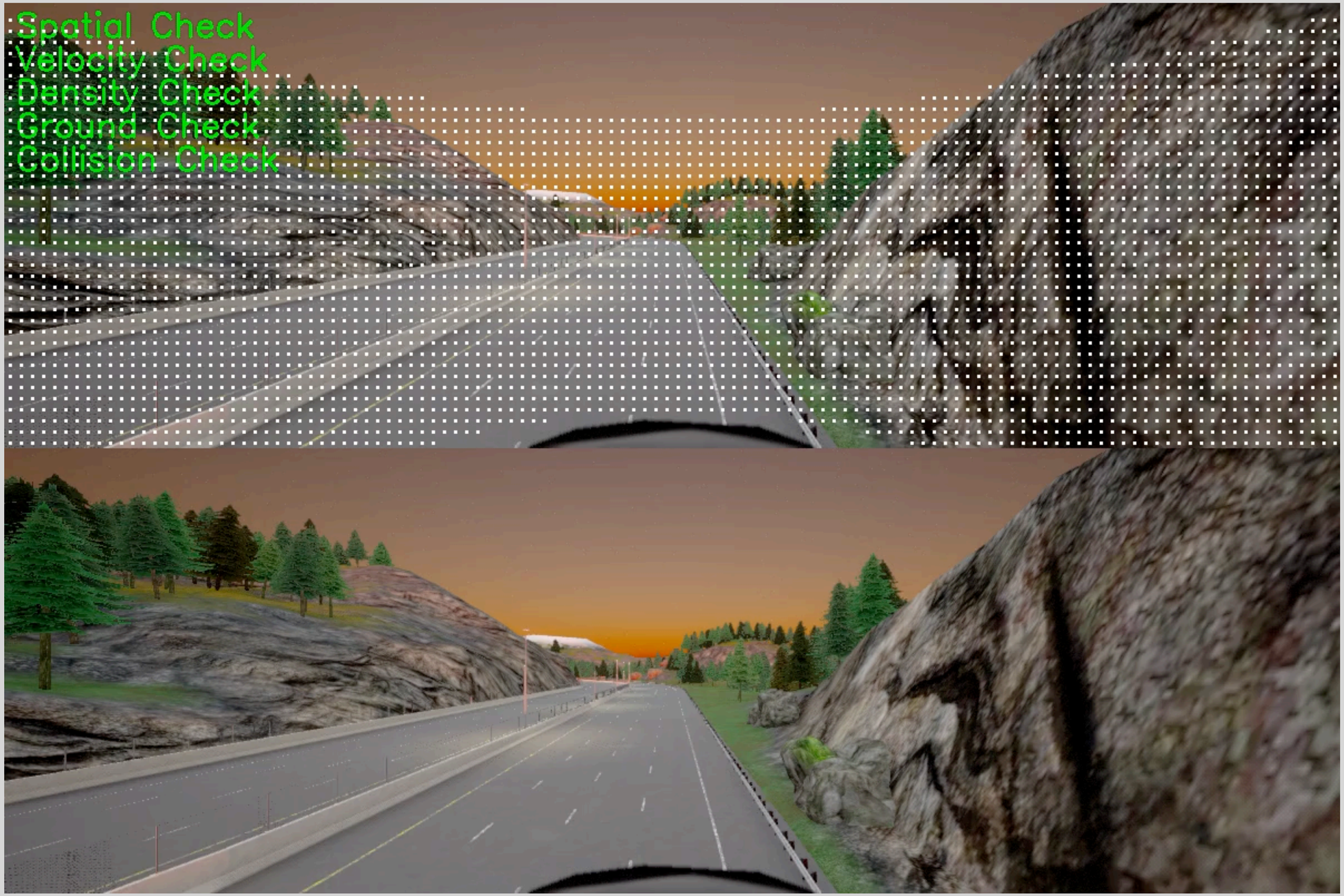
interlock does not interfere when the ADS is safe

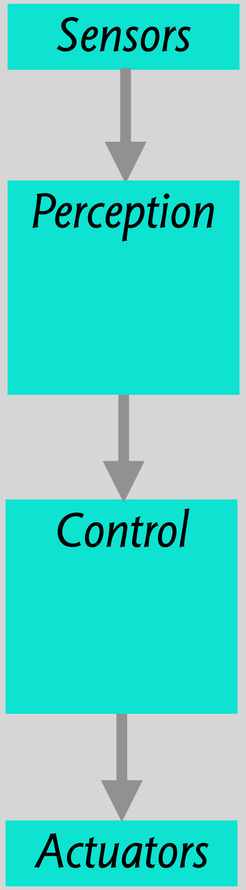


simulating uber tempe accident

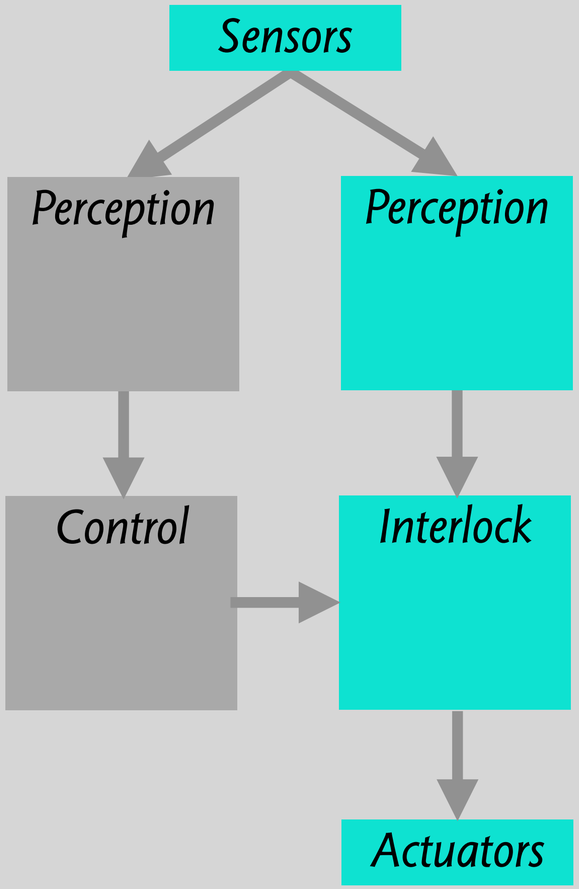


interlock does not interfere when the ADS is safe

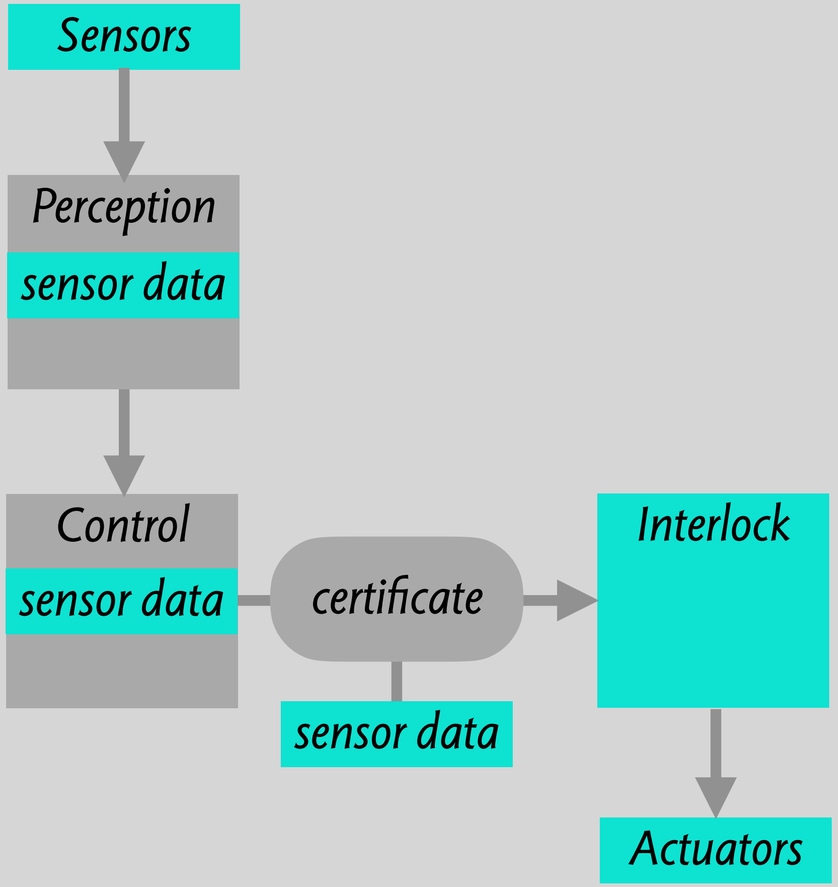




no interlock

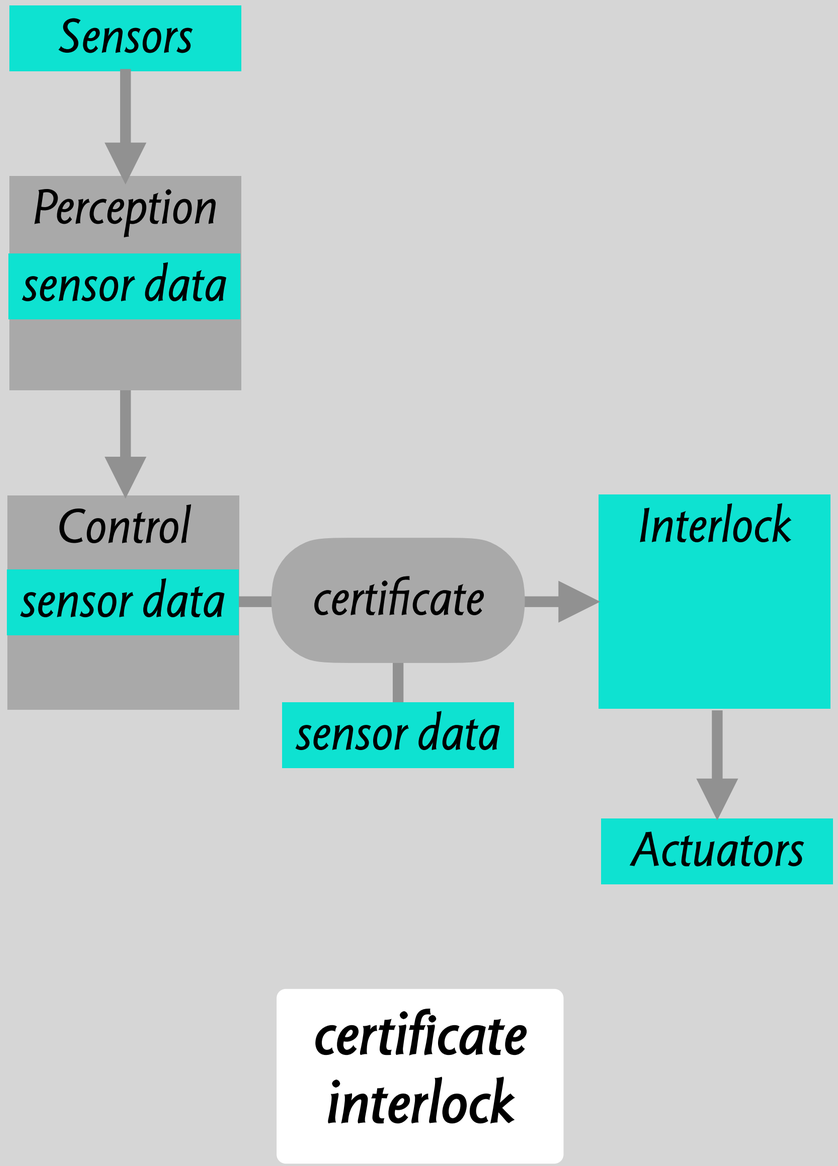


classic interlock



certificate interlock

discussion



what are the risks to certified perception?
machine learning gets so good, we don't need checks
certificate checks are too strong & binary

what are the key challenges?
can't mitigate super-rare events (eg, avalanche)
better at mitigating egregious errors than subtle flaws

what do you think?
will we tolerate "good enough?"
will adversarial attacks matter?