DeepManeuver: Adversarial Test Generation for Trajectory Manipulation of Autonomous Vehicles

Meriel von Stein  David Shriver  Sebastian Elbaum

Problem

Adversarial attacks are often generated for individual images, or for an unordered collection of images meant to represent a test case.

Solution

State-Adaptive Adversarial Testing

Given a DNN for vehicle actuation and a simulator, we build an automated, parameterized framework DeepManeuver that interleaves adversarial test generation with vehicle physics simulation. This creates a refined adversarial test case for the vehicle at the system level.

DeepManeuver

We modify the perturbation generation process to iteratively update the perturbation upon changes in vehicle state to preserve compounding effects of the perturbation in previous states. At each step, new DNN inputs and vehicle actuation values from the simulator are passed to the generator for refinement of the perturbation.

This is enabled through a two-part weighted loss function (below) for the generator that adjusts loss terms according to their importance to the trajectory.

\[
\arg\min_{\eta} \sum_{i=0}^{n-1} w_i L_2(N(\text{img}_i + \eta), \psi_i) + w_0 L_1(N(\text{img}_0 + \eta), \text{target})
\]

Results

Left: Performance comparison of state-of-the-art baseline technique DeepBillboard against DeepManeuver. Success rate denotes fulfillment of the maneuver. Across all parameterizations and environments, DeepManeuver shows a 20.7 percentage point increase in success rate over DeepBillboard.

Right: Performance for three multi-target maneuvers: hit a bullseye, cut a corner, and change lanes.

Bottom: Simplified setup of a scenario to generate a perturbation for a multi-target maneuver.