Maximizing Confidence in Safety-Critical Decisions of Automated Vehicles that are Grounded in ML-based Environmental Perception

— Rendering AVs “Safer than Perception” —

Martin Fränzle[⋆]

Carl von Ossietzky Universität Oldenburg, Germany
martin.fraenzle@uol.de

Keywords: Highly Automated Vehicles, Learning-enabled cyber-physical systems, Environmental perception, AI-based object detection and classification, Safety Assurance.

In this suggested presentation, I will address one of the key challenges in assuring safety of autonomous cyber-physical systems that rely on learning-enabled classification within their environmental perception: How can we achieve confidence in the perception chain, especially when dealing with percepts safe-guarding critical manoeuvres? I will build on (and present) a methodology [1] jointly developed with my colleagues Werner Damm (Oldenburg University, Germany), Willem Hagemann (German Aerospace Center DLR, Institute for Systems Engineering for Future Mobility), Astrid Rakow (also DLR Institute for Systems Engineering for Future Mobility), and Mani Swaminathan (Federal Office for Information Security, Germany) which allows to mathematically prove that the risk of misevaluating a safety-critical guard conditions referring to environmental artefacts can be bounded to a considerably lower frequency than the risk of individual misclassifications, provided that the safety-critical condition is a non-trivial Boolean combination of perceived items. Based on this, I will be presenting a hybrid-AI approach for rigorously optimising true vs. false evaluation rates of such safety-critical conditions and thereby facilitating to adjust misevaluation rates to a value less than a given level of societally accepted risk.

The methodology presented thus is meant to assure the safety of highly automated vehicles (HAV) by guaranteeing that what the ego car, i.e. the own car, believes to be true about its environment and the actual ground truth rarely differ for any aspect relevant for ensuring the safety of the ego vehicle. How rare is rare enough is a matter of societal debates — e.g. the German Department of Transportation requires HAVs to reduce the overall rate of fatalities compared to human-operated vehicles. No matter what order of magnitude these debates will converge to, they will formally result in acceptance thresholds r and s bounding the likelihood of false adoption and of false omission, resp., of

[⋆] Supported by the State of Lower Saxony within the Zukunftslabor Mobilität as well as by Deutsche Forschungsgemeinschaft under grant no. DFG FR 2715/5-1.

safety-critical manoeuvres, where a false positive decision for a safety-critical manoeuvre potentially leads to a risk while a false negative decision against such a manoeuvre potentially incurs a performance penalty. The respective bounds r and s will inevitably be orders of magnitude tighter than what machine-learning-based perception and classification systems can guarantee currently and in the foreseeable future.

In the talk, I set out to bridge this gap between actual perception performance and expected societal acceptance thresholds for safety-critical behaviour by answering the following question: if r (and s, resp.) are the levels of societally accepted risk (and of societally acceptable performance penalty, resp.) budgeted for *relevant* misperceptions induced by false positives (false negatives, resp.) in the evaluation of safety-critical guard conditions, then

1. what renders a misperception "relevant" to a safety-critical decision,
2. how can we mathematically prove that evaluation of guard conditions for safety-critical decisions, being based on perception of relevant environmental artefacts, errs with false-positive rate of at most r while guaranteeing a true-positive rate of at least $1 - s$, and
3. how can we automatically and rigorously transform such safety-critical guard conditions into logically equivalent (within all situations satisfying the invariants of the domain) conditions that feature an optimal trade-off between false-positive and true-positive rates?

We provide a mathematical setting for addressing the above questions 1 and 2 and expose algorithms resolving challenge 3, thereby creating a provably optimal symbolic evaluation layer within a hybrid subsymbolic-symbolic evaluation chain for environmental perception and evaluation of safety-critical decision points.

References

1. Fränzle, M., Hagemann, W., Damm, W., Rakow, A., Swaminathan, M.: Safer than perception: Assuring confidence in safety-critical decisions of automated vehicles. In: Haxthausen, A.E., Huang, W.L.H., Roggenbach, M. (eds.) Applicable Formal Methods for Safe Industrial Products — Essays Dedicated to Jan Pelseka on the Occasion of His 65th Birthday. Lecture Notes in Computer Science, Springer (2023), in print, to appear Aug. 2023