SMILE – Safety analysis and verification/validation of MachIne LEarning based systems

SMILE – Säkerhetsanalys och verifiering/validering av system baserade på maskininlärning

Public report

Author: Cristofer Englund, Markus Borg, Boris Duran

Date: 2017-07-07
Projekt within: Elektronik, Mjukvara och Kommunikation

FFI in short

FFI is a partnership between the Swedish government and automotive industry for joint funding of research, innovation and development concentrating on Climate & Environment and Safety. FFI has R&D activities worth approx. €100 million per year, of which about €40 is governmental funding.

Currently there are five collaboration programs**: Electronics, Software and Communication, Energy and Environment, Traffic Safety and Automated Vehicles, Sustainable Production, Efficient and Connected Transport systems.**

For more information: www.vinnova.se/ffi

1 Summary

This project has explored challenges while introducing machine learning-based systems in automated driving (AD) applications. It also explored strategies to cope with those challenges to guide the industry and thus, be able to realize the potential to apply machine learning in safety critical systems. The project have resulted in a description of a continuation project, SMILE II, that is part of a series of project needed to explore the safety aspects of deep-machine learning while using them in safety-critical vehicle functions.

The project elaborated on the following research questions:

What available methods are there that can guarantee safe use of machine learning algorithms and trained neural networks within safety critical vehicular systems?

Within which safety critical vehicular functions is there a need for machine learning? What are the safety requirements for these systems and which requirements need to be allocated to elements based on machine-learning?

What are the major barriers for introduction of deep machine learning in safety critical systems? What are the available concepts/strategies to improve the level of safety integrity achievable in deep machine learning based systems? Which of these concepts/strategies are worthwhile to develop and evaluate further?

A literature review study based on (a) the initial research questions and (b) discussions with the industrial domain experts within safety critical systems within automation was the first part of the project. The second part concerned a workshop series. Six workshops have been arranged to (i) present the findings from the literature study (ii) give inspirational talks to the project members from researchers outside the project group (iii) cross fertilizing with other research projects, ESPLANADE, (iv) highlight the industrial needs from the project partners, (v) discussions about the industrial needs and their relation to state-of-art, and (vi) concluding the findings from the project and formulate the continuation, SMILE II, project.

A new research agenda is created based on the results from our research. Three main areas were identified: data, models and verification and validation. Data concerns collection and management approaches for training to make robust DNN components. Models is related to data and refers to methods of incorporating data from multiple sources and contexts into the vehicle control systems. The area verification and validation concerns defining test cases, verification/validation environments and procedures, accuracy and safety targets, and other topics related to creating standards for developing DNN-based software used in safety-critical systems.

2 Executive summary

SMILE - Safety analysis and verification/validation of MachIne LEarning based systems, is a concept study with the purpose to explore the challenges while introducing machine learning-based systems in automated driving (AD) applications. It also aims to propose strategies to cope with those challenges to guide the industry and thus, be able to realize the potential to apply machine learning in safety critical systems. The project combined literature reviews and workshop-based discussions with industrial domain experts. The results from these two activities were synthesized into a research agenda for the SMILE research program as well as an application for future research within the SMILE II project.

The agenda for the SMILE program concerns three areas of research, namely data collection and management, deep learning modelling and verification and validation. The underlying concepts are approaches to (a) make robust DNN components, (b) complementing DNN with traditional components and (c) verification & validation approaches for systems with DNN components.

The continuation project will focus on developing a run-time monitoring system for DML-based perception using the concept of *adaptive safety cage architectures* (Heckemann *et al*, 2011), or as referred to by Adler *et al*. (2016): *safety supervisors*. This concept is one of the main results from the SMILE project - where we envision a safety cage that will monitor the input data to the DML-based perception to detect anomalies in the model's classification uncertainty. Varshney *et al*. (2013) describes this as a classifier having a *reject option* when the uncertainty in data and the model is too high, e.g., forcing a human to intervene. In line with the proposal by Heckemann *et al.* (2011), in the next project we will distinguish between a safe region of operation and an invalid region that could lead to a dangerous situation. If the DML-based perception enters the invalid region, the safety cage will invoke an appropriate safe action, such as graceful degradation based on deterministic algorithms. The design of the safe action is planned for the next stage of the SMILE research program (SMILE III).

3 Background

There is currently considerable momentum in the area of autonomous vehicles (Knauss *et al.,* 2017). One critical enabler for the technology has been the application of DML for perception, mainly relying on computer vision and forward looking cameras in vehicles. Several studies show that DML can successfully handle complex traffic data (Huval *et al*., 2015) (LeCun *et al*., 2015). However, no machine learning model will be sufficiently complete to avoid misbehavior under all circumstances on the road (Spanfeiner *et al*., 2012), thus also DML will sometimes fail to generalize. Unfortunately, the models trained using DML are particularly opaque in nature, as they often consist of huge networks with a number of parameter weights in the order of magnitude of hundreds of millions (Han *et al*., 2016). Consequently, there are very limited options to analyze miss-classifications from a functional safety perspective, as neither traditional code reviews nor exhaustive safety analysis techniques are possible.

Members of the SMILE consortium are active in both national and international research projects, considerably increasing the potential dissemination and impact of any novel research findings. One of the closest related projects is ESPLANADE, coordinated by RISE SP. ESPLANADE is a three-year project with several partners that aims at creating a common safety methodology for autonomous vehicles. As the interests of SMILE and ESPLANADE are clearly overlapping, we have already organized a common workshop with all 13 partners present, in total 30 persons - we anticipate further collaboration also during SMILE II.

Viktoria is active in several related projects, for example, Viktoria is leading the Vehicle ICT Arena (VICTA) Lab where the electrical architecture of a vehicle is simulated. Current research includes incorporating validated sensors to be able to make realistic simulations also with external sensors such as cameras, radar and lidar. This is an enabler of the SMILE program that can allow efficient testing of the developed software within VICTA Lab. Other related projects including machine learning includes AIR, where Viktoria develops algorithms for both predicting vehicle behaviour based on ML algorithms as well as face expression detection based on DML trained on camera data. QRTECH currently leads the demonstrator work package in the EU project TRACE. The demonstrator uses lidar, radar and stereo camera images processed through deep neural networks to implement DML-perception and object identification for autonomous cars. Currently, the system does not apply a safety cage concept. In parallel, QRTECH participates in the newly initiated AutoDrive, which will look into architectural measures for increasing confidence in neural networks. Dr. Markus Borg at RISE SICS is WP leader in the ITEA3 project TESTOMAT on test automation. Among other goals, TESTOMAT will study automated testing targeting non-functional system properties such as functional safety, i.e., research that might also be highly relevant to the SMILE program. The Swedish TESTOMAT consortium includes Bombardier, Saab, Ericsson and several SME. Finally, if projects within the SMILE program at some point requires considerable computational resources, RISE SICS operates a datacenter in Luleå that is available for any task that requires infrastructure of big data magnitude.

4 Purpose, research questions and method

The purpose of SMILE is to explore the challenges while introducing machine learning-based systems in automated driving (AD) applications. It also aims to propose strategies to cope with those challenges to guide the industry and thus, be able to realize the potential to apply machine learning in safety critical systems.

This is achieved by elaborating on the following research questions:

What available methods are there that can guarantee safety within machine learning algorithms and trained neural networks that are applied within safety critical vehicular systems?

Within which vehicular systems is there a need for machine learning? and What safety requirements do these systems have?

What are the major barriers for introduction of deep machine learning in safety critical systems? How can new paths forward be created and what future concepts should be developed and evaluated to show that the safety is achieved in deep machine learning based safety critical systems?

The methods that have been used in SMILE to answer these questions is a literature review study based on (a) the initial questions and (b) discussions with the industrial domain experts within safety critical systems within automation. The second method used is workshop and discussion-based. Five workshops have been arranged to (i) present the findings from the literature study (ii) give inspirational talks to the project members from researchers outside the project group (iii) cross fertilizing with other research projects, ESPLANADE, (iv) highlight the industrial needs from the project partners and (v) concluding the findings from the project and formulate the continuation, SMILE II, project. The method is illustrated in Fig. 1.

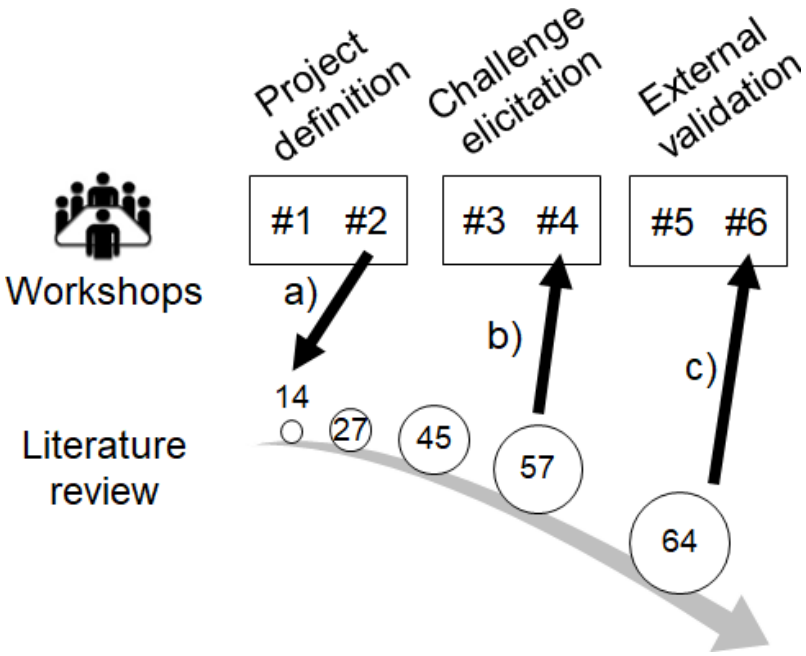

Fig. 1 Illustration of the proposed method used in SMILE.

5 Objective

The purpose of SMILE is to explore the challenges while introducing machine learning-based systems in automated driving (AD) applications. It also aims to propose strategies to cope with those challenges to guide the industry and thus, be able to realize the potential to apply machine learning in safety critical systems.

6 Result and deliverables

The FFI goals and their applicability in the SMILE project:

- Increasing the Swedish capacity for research and innovation, thereby ensuring competitiveness and jobs in the field of vehicle industry
    - This goal is addressed and great interest in shown for the continuation project from current and new partners.
- Developing internationally interconnected and competitive research and innovation environments in Sweden
    - The specific interest from other projects in this domain confirms the relevance of the SMILE, and SMILE II project. The dissemination activities, poster presentations, visits to conferences etc. also report on a large interest in our research.
- Promoting the participation of small and medium-sized companies
    - The large involvement of QRTECH, an SME with expert knowledge within deep learning, has given the project energy and highlights the needs of this research.
- Promoting the participation of subcontractors
    - QRTECH is a subcontractor already and the great interest from other subcontractors that are part of the ESPLANADE project e.g. Semcon, Qamcom also highlights the interest within this group of actors.
- Promoting cross-industrial cooperation
    - The current state-of-art within this field is mainly driven by automation in general, the closest research is found from the avionics field.
- Promoting cooperation between industry, universities and higher education institutions
    - Two master thesis projects are executed in relation to this project, one at Blekinge Tekniska Högskola, co-supervised by SICS, concerning a snowballing approach to a literature review about machine learning in safety-critical systems. The second master thesis project that is related to SMILE is at Chalmers, co-supervised by VIKTORIA, regarding end-to-end reinforcement learning for an autonomous vehicle.
- Increase the technical maturity level (by measuring “technology readiness level” (TRL) and rationalize product development methods in order to achieve faster time-to-market and increased customer value
    - SMILE is a pre-study that successfully extracted the industrial needs, in relation to the state-of-art and formulated and submitted a continuation project aiming at developing and testing new software that can allow the application of deep machine learning in safety critical vehicular systems. If SMILE II is granted, it will work towards lifting this technology from TRL 2 to level 4.

As a result of the concept study SMILE, a research agenda for a research program called SMILE was created and an application for a continuation project, named SMILE II, was created and submitted to FFI - Maskininlärning on June 13 2017. The project also resulted in

an application for an Institute PhD student where funding was applied at SSF - Stiftelsen för Strategisk Forskning.

While SMILE studied the state-of-art and industrial needs. SMILE II, will aim at developing run-time monitoring for DML-based perception using the concept of *adaptive safety cage architectures* (Heckemann *et al*, 2011), or as referred to by Adler *et al*. (2016): *safety supervisors*. This concept is one of the main results from the SMILE project - where we envision a safety cage that will monitor the input to the DML-based perception to detect anomalies in the model's classification uncertainty. Varshney *et al*. (2013) describes this as a classifier having a *reject option* when the uncertainty is too high, e.g., forcing a human to intervene. In line with the proposal by Heckemann *et al.* (2011), in the next project we will distinguish between a safe region of operation and an invalid region that could lead to a dangerous situation. If the DML-based perception enters the invalid region, the safety cage will invoke an appropriate safe action, such as graceful degradation based on deterministic algorithms. The design of the safe action is planned for the next stage of the SMILE research program (SMILE III).

The motivation for focusing on safety cages is that current alternatives to prevent system failures, e.g., fault prevention and fault avoidance, cannot address all malfunctions due to the complexity of the system and the non-deterministic traffic environment (Ramos *et al.* 2017). Instead, our long-term goal is to accomplish ASIL decomposition by developing the safety cage as a functionally redundant system to the actual control system. For such a solution, the complex control function (i.e., applying DML) could be developed according to the quality management standard, whereas the comparably simple and deterministic safety cage could be addressed by traditional verification and validation (V&V), or possibly even proven correct using formal verification methods (Abdulkhaleq *et al*., 2015).

A useful by-product from applying run-time monitoring to detect anomalies (i.e, the envisioned safety cage) is the possibility to collect a set of images for which the DML-based perception is the least certain. Such a dataset could later be used to further increase the robustness of the DML-models, both by supporting interpretability of classification results and by guiding future collection of training data. Guidance of training data collection is analogous to the concept of uncertainty sampling in *active learning* (Settles, 2012), i.e., enabling a DML model to perform better with less training by actively selecting training data. Moreover, the intention is to evaluate model-agnostic methods to trace back from specific DML miss-classifications (and highly uncertain examples) to particularly influential features in the data, such as recently proposed for image recognition by Ribeiro *et al.* (2016). Finally, we plan to explore using the dataset of highly uncertain images to update DML models with the new data using approaches developed for single-view instance recognition (Held *et al.*, 2015).

7 Dissemination and publications

7.1 Dissemination

Hur har/planeras projektresultatet att användas och spridas?	Markera med X	Kommentar
Öka kunskapen inom området	**x**	
Föras vidare till andra avancerade tekniska utvecklingsprojekt	**x**	
Föras vidare till produktutvecklings- projekt	**x**	
Introduceras på marknaden		
Användas i utredningar/regelverk/ tillståndsärenden/ politiska beslut		

The SMILE project has been presented at the following events: the workshop Grand Challenges of Traceability: The Next Ten Years in Slade, Kentucky, US and the International Summer School on Deep Learning in Bilbao, Spain (Fig 2).
Furthermore, SMILE has presented during poster presentations at The Digital Society Symposium during Lund University 350th Anniversary, Lund, Sweden, April 24-25, 2017 and at SICS Open House, Kista, Sweden, May 17, 2017

Fig. 2 Picture from presentation at the summer school in deep learning in Bilbao 2017.

As a result, from SMILE, an application to SSF about a Institute PhD student is filed about Deep Learning in safety-critical systems. If granted, the project will start in the fourth quarter of 2017. A use-case where new cars are being driven onboard RoRo vessels is highlighted, the task is, among other things, to identify vehicles as they enter a RoRo vessel.

7.2 Publikationer

C. Englund, M. Borg, B. Duran, H. Kaijser, H. Lönn, K. Lindström, C. Zandén, C. Lewandowski, M. Zymon, J. Törnquist. Deep Learning and Safety-critical Systems: Research, Practice, and Future Needs in Automotive, Draft paper intended for submission to *Transactions on Intelligent Transportation Systems*

M. Borg, C. Englund, and B. Duran. Traceability and Deep Learning – Safety-critical Systems with Traces Ending in Deep Neural Networks, In *Proc. of Grand Challenges of Traceability: The Next Ten Years*, Slade, Kentucky, USA, 2017.

8 Conclusion and future research

The SMILE project have extracted, presented and discussed approaches to engineer robust DNN components in order to increase reliability of perception (object detection and classification) with verification & validation of automotive perception based on DNN to allow safety certification.

A similar study as the one brought out in the SMILE project, by US Air Force Clark et.al (2014), found four enduring problems in relation to certification of autonomous systems:

(i) State-Space Explosion,
(ii) Unpredictable Environments,
(iii) Emergent Behavior,
(iv) Human-Machine Communication,

that, according to the findings in our research, also characterize the field of test and evaluation of Deep learning-based safety-critical autonomous vehicular systems.
In conjunction to the challenges, goals for future research was also put forward.

The alignment of those goals, namely

(v) Cumulative Evidence through Research & Development,
(vi) Test & Evaluation and operational tests,
(vii) Evidence generated during design,
(viii) Requirements Development and Analysis,
(ix) Decision Assurance, Compositional Case Generation

and the automotive industrial needs (found during the workshop series), approaches to make robust DNN components, complementing DNN with traditional components and V&V approaches for systems with DNN components, elicited in our research is clear. In addition, they also lay the foundation to our future research agenda.

A new research agenda is created based on the results from our research. Three main areas were identified namely, data, modelling and V&V in accordance to the findings from the workshop series. More specifically, approaches to make robust DNN components concerns to a large extent data collection and management. The modelling area is also related to data, in particular, methods such as transfer learning are expected to be an essential part of future research to be able to incorporate data from different contexts and sources into the vehicle control systems. The final area concerns V&V methods that can be used to define test cases, accuracy and safety targets, to create standards for developing DNN-based software used in safety-critical systems.

9 Participating parties and contact persons

Cristofer Englund, RISE Viktoria
cristofer.englund@ri.se
0708 560 227

Jacob Axelsson, RISE SICS
jacob.axelsson@ri.se
072 734 29 52

Lars-Åke Johansson, QRTECH
lars-ake.johansson@qrtech.com
070 824 60 30

Henrik Kaijser, Volvo Technology
henrik.kaijser@volvo.com
073 9025774

Erik Hjerpe, Volvo Car Group
erik.hjerpe@volvocars.com
031-59 64 56

References

K. Heckemann, M. Gesell, T. Pfister, K. Berns, K. Schneider, and M. Trapp, “Safe Automotive Software,” in Knowledge-Based and Intelligent Information and Engineering Systems. Springer, Berlin, Heidelberg, 2011, pp. 167–176.

K. R. Varshney, R. J. Prenger, T. L. Marlatt, B. Y. Chen, and W. G. Hanley, “Practical ensemble classification error bounds for different operating points,” IEEE Trans. on Knowl. and Data Eng., vol. 25, no. 11, pp. 2590–2601, Nov. 2013. [Online]. Available: http://dx.doi.org.ludwig.lub.lu.se/10.1109/TKDE.2012.219

A. Knauss, J. Schroeder, C. Berger, and H. Eriksson, “Software-related challenges of testing automated vehicles,” in Proceedings of the 39th International Conference on Software Engineering Companion, ser. ICSE-C ’17. Piscataway, NJ, USA: IEEE Press, 2017, pp. 328–330. [Online]. Available: https://doi-org.ludwig.lub.lu.se/10.1109/ICSEC.2017.67

Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 5 2015.

B. Spanfeiner, D. Richter, S. Ebel, U. Wilhelm, W. Branz, and C. Patz, “Challenges in applying the ISO 26262 for driver assistance,” in Proc. of the Schwerpunkt Vernetzung, 5. Tagung Fahrerassistenz, Munich, Germany, 2012.

Adler, Feth, and Schneider. Safety Engineering for Autonomous Vehicles, In Proc. of the 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops, pp. 200-205, 2016.

Settles, B. Active Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning, 6(1), pp. 1-114, Morgan & Claypool Publishers, 2012.

Ramos, Gehrig, Pinggera, Franke, and Rother. Detecting Unexpected Obstacles for Self-Driving Cars: Fusing Deep Learning and Geometric Modeling, In Proc. of the IEEE International Conference on Robotics and Automation, 2017.

Abdulkhaleq, Wagner, and Leveson. A Comprehensive Safety Engineering Approach for Software- Intensive Systems Based on STPA, Procedia Engineering, 128, pp. 2-11, 2015.

Settles, B. Active Learning, Synthesis Lectures on Artificial Intelligence and Machine Learning, 6(1), pp. 1-114, Morgan & Claypool Publishers, 2012.

Ribeiro, Singh, and Guestrin. Why Should I Trust You? Explaining the Predictions of Any Classifier, In Proc. of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1135- 144, 2016.

M. Clark, K. Kearns, J. Overholt, K. Gross, B. Barthelemy, and C. Reed, “Air force research laboratory test and evaluation, verification and validation of autonomous systems – Challenge

exploration final report," Air Force Research Laboratory, Wright-Patterson AFB, OH, US, Tech. Rep., 2014.