Towards Provable Safety in Lane Detection and Monitoring for Autonomous Vehicles

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems

Background

Although autonomous systems are fast becoming prevalent in a wide variety of use-cases, they are currently an especially relevant topic in the transportation sector. Technologies such as autonomous driving and fully autonomous railway traffic have the potential to revolutionize the concept of personal mobility, lower costs, and improve safety. Considerable progress has been made recently in machine perception, such as object detection and scene understanding, using neural networks, thus improving the reliability of these autonomous systems [8, 11, 6]. However, in such safety-critical use-cases, full automation has to be predicated on the provable safety of detection algorithms responsible for recognizing and classifying scenarios during operation.

A key challenge on the path to safe deployment is the susceptibility of neural networks to adversarial attacks, sensor noise, and other disturbances [3]. Such perturbations, though small, can profoundly influence model outputs and consequently cause hazardous system behaviour with potentially catastrophic consequences. Furthermore, verifying image-based neural networks in real-time is computationally challenging due to the high-dimensional continuous input space [7].

Formal verification methods, which provide mathematical proofs of correct behavior, offer a promising avenue for certifying safety [10, 2]. Recent research on set-based verification, such as using zonotopes and abstract interpretation techniques, has shown promise in analyzing deep neural networks in a scalable way [9]. However, most existing approaches have not yet been extended to tackle the verification of networks for entire video sequences. Additionally, they usually only analyze the robustness of neural networks in a local neighborhood around some input [2].

This thesis work aims to develop a flexible approach to provably safe lane detection and monitoring for certain types of autonomous vehicles. To this end, it aims to improve a recently developed novel framework for safe railway track monitoring, and extend it to the use-case of provably safe lane detection and monitoring for autonomous cars.

Description

The original approach to safe railway monitoring partitions the input space based on safety specifications in the network's output space, thus allowing the online verification of new images via simple containment checks. The goal of this thesis is to extend this approach to a new domain, provably safe lane detection for autonomous vehicles. The existing pipeline can be generalized and adapted to the new domain by exploiting similarities in the scenarios, although several unique challenges must also be addressed.

In order to overcome the computational challenges associated with verifying image-based neural networks, the original approach developed a dimensionality reduction pipeline tailored to the railway use-case. This pipeline automatically segments out safety-critical image regions based on ground truth annotations, thus reducing the dimensionality of the problem. In a first step, the functionality of this pipeline must be adapted to the automotive context. Furthermore, to fully automate the pipeline for online deployment, a necessity for autonomous vehicles, it must be improved so that segmentation relies only on readily available sensor and map data.

Despite promising results, the original approach still suffers from a series of restrictive assumptions and algorithmic limitations which limit its performance. Specifically, while the algorithm successfully eliminates the risk of catastrophic False-Negative predictions (i.e., the situation is unsafe but classified as safe), it does so at the cost of a large number of False-Positives due to it's overly conservative and naive partitioning of the input space. Furthermore, it assumes a correct neural network (i.e., one that does not misclassify any samples), which limits the real-world applicability of the scheme. Improving such shortfalls to enable robust and provably safe online lane detection represents another significant portion of this work.

Supervisor:
Prof. Dr.-Ing. Matthias Althoff

Advisor:
Tobias Ladner, M.Sc.

Research project:
-

Type:
MT

Research area:
Formal verification, neural networks

Programming language:
MATLAB, Python

Required skills:
Knowledge in formal methods and machine learning, good mathematical background

Language:
English

Date of submission:
29. Oktober 2025

For more information please contact us:

Phone: +49 (89) 289 - 18140

E-Mail: tobias.ladner@tum.de

Website: ce.cit.tum.de/cps/

Tasks

- Literature research of relevant work related to the task of image/video verification for neural networks.
- Extend the dimensionality reduction pipeline to lane detection in the automotive use-case, thus allowing the application of the developed verification approach.
- Modify the dimensionality reduction pipeline to allow fully automatic segmentation of safety critical image regions based only on GNSS and camera data, as well as map information. Test the functionality using a dataset collected with the EDGAR test vehicle.
- Improve the heuristics used for input splitting during the refinement of the input space, for example, by using the set-based gradient to guide splitting, in order to reduce the conservativeness of the algorithm.
- Modify the algorithm to take full advantage of the reduced dimensionality of the input space and return unsafe regions in the form of constrained zonotopes instead of intervals [4], thus also reducing the conservativeness of the computation.
- Mitigate the restrictiveness of assuming a correct neural network, either by simplifying the problem and employing a simple, explainable predictor, or, by using novel methods such as set-based training [5] to increase the robustness of the network's decision boundaries, thus making the assumption more tenable.
- Evaluate the online performance of the approach in both use-cases using unseen test sequences with synthetically generated unsafe artifacts.

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems

References

[1] Matthias Althoff. An introduction to CORA 2015. In *ARCH@ CPSWeek*, pages 120–151, 2015.

[2] Christopher Brix, Stanley Bak, Taylor T Johnson, and Haoze Wu. The fifth international verification of neural networks competition (vnn-comp 2024): Summary and results. *arXiv preprint arXiv:2412.19985*, 2024.

[3] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations*, 2015.

[4] Lukas Koller, Tobias Ladner, and Matthias Althoff. Out of the shadows: Exploring a latent space for neural network verification. *arXiv preprint arxiv:2505.17854*, 2025.

[5] Lukas Koller, Tobias Ladner, and Matthias Althoff. Set-based training for neural network verification. *Transactions on Machine Learning Research (TMLR)*, 2025.

[6] Quoc-Vinh Lai-Dang. A survey of vision transformers in autonomous driving: Current trends and future directions. *arXiv preprint arxiv:2403.07542*, 2024.

[7] Aditya Parameshwaran and Yue Wang. Scalable and interpretable verification of image-based neural network controllers for autonomous vehicles. *arXiv preprint arxiv:2501.14009*, 2025.

[8] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[9] Gagandeep Singh, Timon Gehr, Matthew Mirman, Markus Püschel, and Martin Vechev. Fast and effective robustness certification. *Advances in Neural Information Processing Systems*, 31, 2018.

[10] Caterina Urban and Antoine Miné. A review of formal methods applied to machine learning. *arXiv preprint arXiv:2104.02466*, 2021.

[11] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017.