Analysis of Neural Activation Patterns During Video Segmentation for Formal Verification

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems

Background

Neural networks are increasingly used in safety-critical environments such as autonomous driving [2, 4]. However, they require formal verification due to their vulnerability to adversarial attacks [6]. This verification is crucial, as erroneous outputs can lead to catastrophic results. Formal verification methods provide a mathematical guarantee of a neural network's behavior, ensuring it operates as expected in various conditions and maintains safety standards in these environments.

For instance, autonomous trains use neural networks to inspect their environment [10]. This includes tasks like railway track segmentation [3], where accuracy is crucial for detecting obstacles and ensuring safe navigation. Ensuring that neural networks used in autonomous trains operate reliably requires robust verification techniques. Set-based formal verification [8, 9] uses representations such as zonotopes [5] and polynomial zonotopes [7] to create outer-approximations of the network's outputs. This method ensures that the network's predictions are enclosed within safe bounds.

Description

The goal of this thesis is to develop a set-based formal verification approach for neural networks used in binary segmentation of railway tracks. The OSDaR23 dataset [11], which includes annotated images of railway tracks under diverse conditions, offers a good foundation for training and evaluating these networks. The methodology will start by training a neural network using the OSDaR23 dataset to classify pixels as railway track or background. After that, we use the CORA framework to verify the neural network's behavior and eventually extend the verification to sequences of frames to analyze changes in neural activation patterns over time, in order to optimize verification for continuous video data.

Tasks

- Literature research on set-based neural network verification
- Familiarize with the CORA toolbox [1]
- Dataset selection and preprocessing
- Implementation and training of a neural network for binary segmentation of railway tracks
- Evaluation of activation patterns
- Optional: Verification of frame sequences to analyze the network's behavior over time and optimize the verification process for sequential input data

References

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

[2] Chih-Hong Cheng, Frederik Diehl, Yassine Hamza, Gereon Hinz, Georg Nührenberg, Markus Rickert, Harald Ruess, and Michael Troung-Le. Neural networks for safety-critical applications - challenges, experiments and perspectives, 09 2017.

[3] Chen Chenglin, Wang Fei, Yang Min, Qin Yong, and Bai Yun. Edge-enabled real-time railway track segmentation, 2024.

[4] Pranav Singh Chib and Pravendra Singh. Recent advancements in end-to-end autonomous driving using deep learning: A survey, 2023.

[5] Antoine Girard. Reachability of uncertain linear systems using zonotopes. In *Proceedings of the 8th International Conference on Hybrid Systems: Computation and Control*, HSCC'05, page 291–305, Berlin, Heidelberg, 2005. Springer-Verlag.

Supervisor:
Prof. Dr.-Ing. Matthias Althoff

Advisor:
Tobias Ladner, M.Sc.

Research project:
FAI

Type:
BT

Research area:
Formal verification, neural networks

Programming language:
Python, MATLAB

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

Language:
English

Date of submission:
21. Juni 2024

For more information please contact us:

Phone: +49 (89) 289 - 18140

E-Mail: tobias.ladner@tum.de

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

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

[7] Niklas Kochdumper and Matthias Althoff. Sparse polynomial zonotopes: A novel set representation for reachability analysis. *IEEE Transactions on Automatic Control*, 66(9):4043–4058, September 2021.

[8] Niklas Kochdumper, Christian Schilling, Matthias Althoff, and Stanley Bak. *Open- and Closed-Loop Neural Network Verification Using Polynomial Zonotopes*, page 16–36. Springer Nature Switzerland, 2023.

[9] Tobias Ladner and Matthias Althoff. Automatic abstraction refinement in neural network verification using sensitivity analysis. In *Proceedings of the 26th ACM International Conference on Hybrid Systems: Computation and Control*. ACM, 2023.

[10] Prashant Singh, Maxim A. Dulebenets, Junayed Pasha, Ernesto D. R. Santibanez Gonzalez, Yui-Yip Lau, and Raphael Kampmann. Deployment of autonomous trains in rail transportation: Current trends and existing challenges. *IEEE Access*, 9:91427–91461, 2021.

[11] Rustam Tagiew, Pavel Klasek, Roman Tilly, Martin Köppel, Patrick Denzler, Philipp Neumaier, Tobias Klockau, Martin Boekhoff, and Karsten Schwalbe. Osdar23: Open sensor data for rail 2023. In *2023 8th International Conference on Robotics and Automation Engineering (ICRAE)*. IEEE, November 2023.

Technical University of Munich

Department of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems