Search-Based Robustness Testing for Deep Learning Computer Vision Systems

Bachelor's Thesis, Master's Thesis

Supervisor: Prof. Dr. Alexander Pretschner
Advisor: Simon Speth
Email: simon.speth@tum.de
Phone: +49 (89) 289 - 17836
Starting date: Any time

Fakultät für Informatik
Lehrstuhl 4
Software & Systems Engineering
Prof. Dr. Alexander Pretschner

Boltzmannstraße 3
85748 Garching bei München

Tel: +49 (89) 289 - 17836
https://www4.in.tum.de

Context

Today, most OEMs equip their state-of-the-art vehicles with traffic sign recognition systems. In those vehicles the traffic sign recognition is performed by computer vision systems, which are usually implemented with Deep Neural Networks (DNNs).

Unfortunately, it has been shown many times that even state-of-the-art DNNs can easily be fooled by adversarial attacks [1]. It is even more concerning that such attacks do not only work in lab settings but also in realistic settings [2, 3, 4]. In particular [5] vividly demonstrate this by placing stickers on real traffic signs.

Goal

The goal of this thesis is to implement a search-based test case generation for DNNs and evaluate a metric of how robust a deep learning computer vision system is against a certain physical attack method.

Based on the obtained results, one can make an argumentation that a system tested in a tough situation will work flawless in other less critical situations. Finally, this makes a contribution towards verification and correctness of machine learning systems and how search can help finding "good" test cases [6] which reveals potential defects of a machine learned system.

Working Plan

1. Familiarize yourself with the literature on testing machine learning systems
2. Conduct a literature survey on physical world adversarial attacks
3. Write the exposé
4. Search for a state-of-the-art pre-trained traffic sign detection deep learning model and dataset
5. Implement the search-based test case generation
   (a) Formulate the search space and implement the image generation for this exact search problem
   (b) Perform the search-based test case generation (implement the actual search)
   (c) Derive robustness metrics (e.g. the minimum area, number, and position of stickers placed on the traffic sign)
   (d) Evaluate the implemented approach. This is a two-fold task. On the one hand, the implemented test case generation should be evaluated. On the other hand, the underlying deep learning model should be evaluated on the basis of the derived robustness metrics.
6. Write the thesis report

Deliverables

- Exposé (about 6 weeks after kick-off)
- Source code of the implementation.
- Short technical report with comprehensive documentation of the implementation, i.e. design decision, technologies used and usage instructions.
- Final thesis report written in English and in conformance with TUM guidelines
- Presentation of the work at the chair (2-3 weeks after submission)

References

[1] Naveed Akhtar and Ajmal Mian. "Threat of adversarial attacks on deep learning in computer vision: A survey". In: *Ieee Access* 6 (2018), pp. 14410–14430.

[2] Husheng Zhou et al. “Deepbillboard: Systematic physical-world testing of autonomous driving systems”. In: *2020 IEEE/ACM 42nd International Conference on Software Engineering (ICSE)*. IEEE. 2020, pp. 347–358.

[3] Zelun Kong et al. “Physgan: Generating physical-world-resilient adversarial examples for autonomous driving”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 14254–14263.

[4] Kaichen Yang et al. “Beyond Digital Domain: Fooling Deep Learning Based Recognition System in Physical World”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 01. 2020, pp. 1088–1095.

[5] Kevin Eykholt et al. “Robust physical-world attacks on deep learning visual classification”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pp. 1625–1634.

[6] Alexander Pretschner. “Defect-Based Testing.” In: *Dependable Software Systems Engineering* 84 (2015).

Fakultät für Informatik
Lehrstuhl 4
Software & Systems Engineering
Prof. Dr. Alexander Pretschner

Boltzmannstraße 3
85748 Garching bei München

Tel: +49 (89) 289 - 17836
https://www4.in.tum.de