Verification and Validation for Machine Learning

Software Verification and Validation Homework, 2020.12.10.

Presenter: István Fábián

Presented paper: https://arxiv.org/abs/1812.05389

BME AUT
Department of Automation and Applied Informatics

Presented Paper

- Safely Entering the Deep: A Review of Verification and Validation for Machine Learning and a Challenge Elicitation in the Automotive Industry
- Motivation:
  - Autonomous driving + ML → hot topic

https://commons.wikimedia.org/wiki/File:Self_driving_Uber_prototype_in_San_Francis

Introduction – ISO 26262

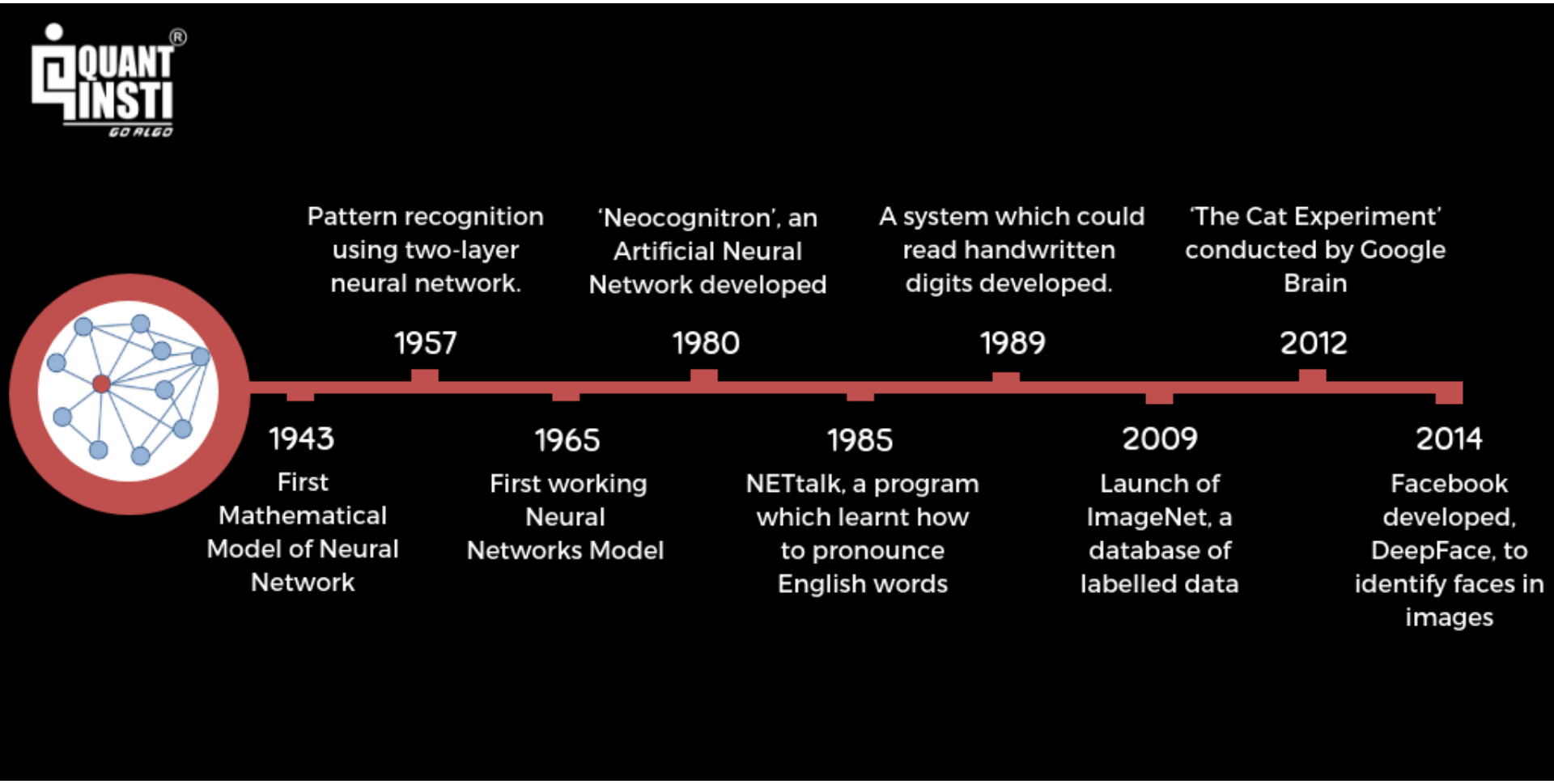
- "Road vehicles – Functional safety“
  - > Adaptation of the Functional Safety standard IEC 61508 for Automotive Electric/Electronic Systems
  - > 1st edition (2011): passenger cars < 3500 kg
  - > 2nd edition (2018): all road vehicles (except mopeds)
- Risk-based safety standard
  - > Assess hazardous operational situations
  - > Avoid, control or detect systematic failures or mitigate their effects
- Problem: can’t handle the major paradigm shift by ML

Introduction – Deep Learning

- Role of DL in autonomous driving:
  - > Classify camera data
  - > Environmental perception, awareness of elements
  - > Lane departure detection, path planning, vehicle tracking
  - > Etc.

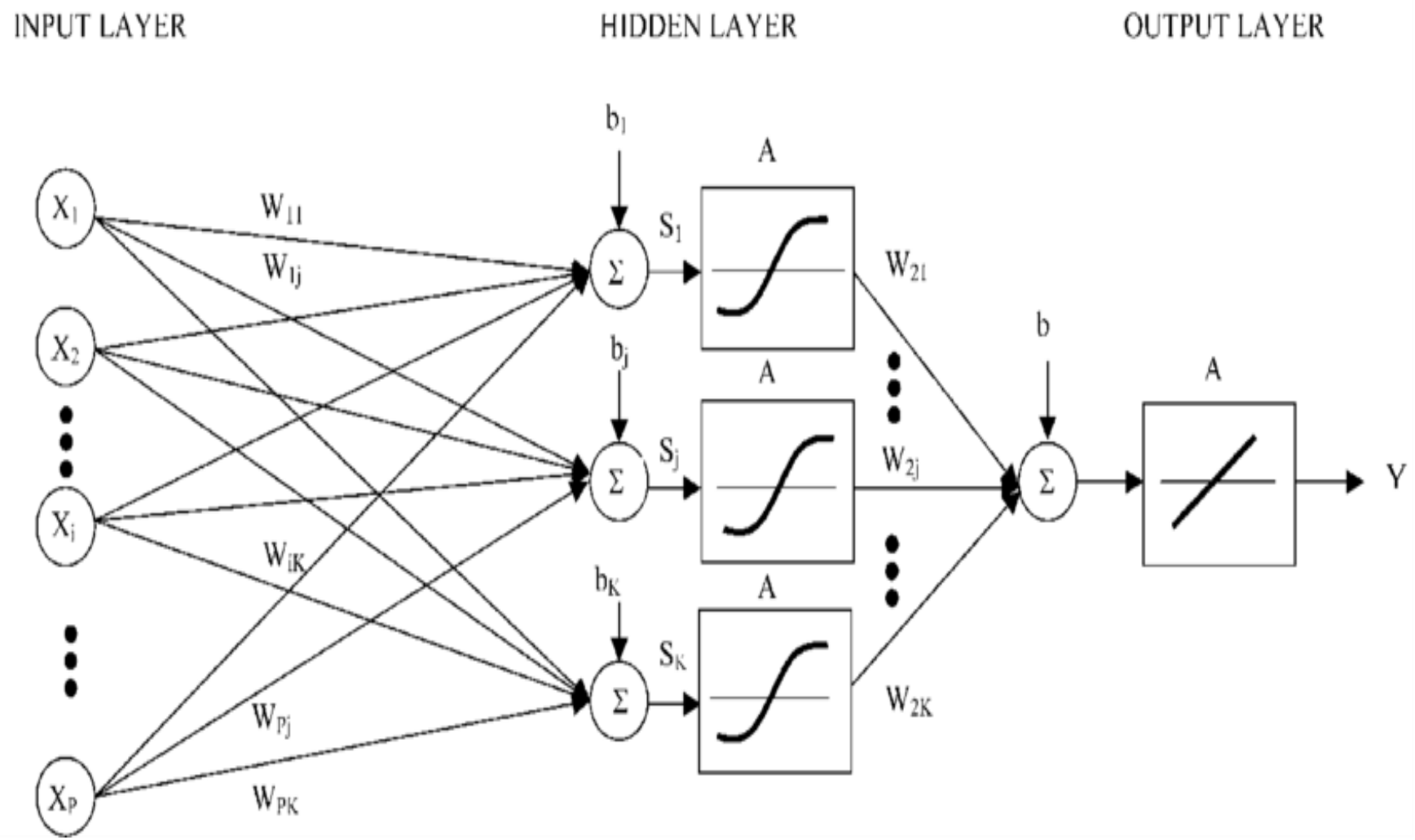
https://blogs.nvidia.com/blog/2019/08/21/drive-labs-autonomous-vehicle-rid

Introduction – Deep Learning



https://blog.quantinsti.com/introduction-deep-learning-neural-netv

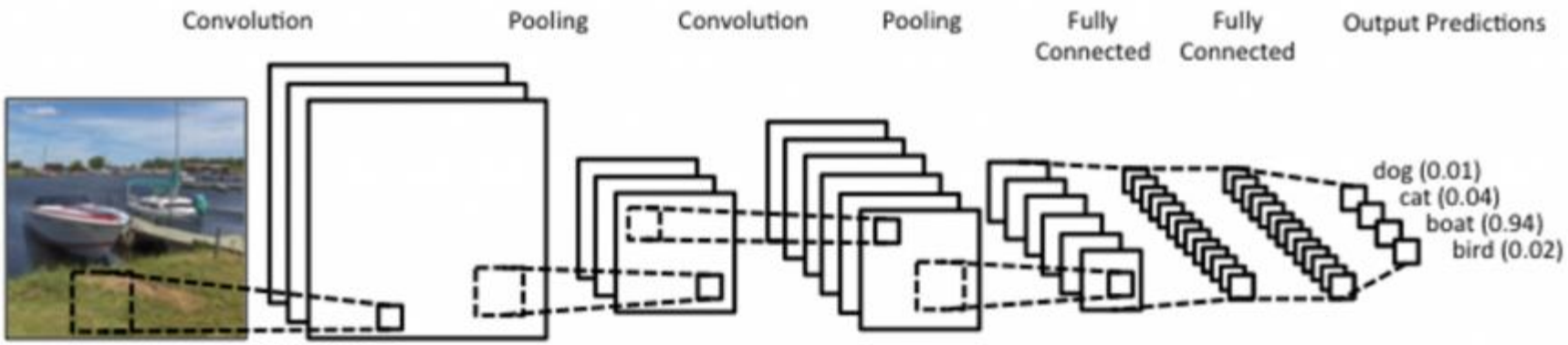
Introduction – Deep Learning



https://www.researchgate.net/figure/Feed-forward-neural-network-with-sigmoid-activation-function-X-i-i-1P-input_fig2_273204474

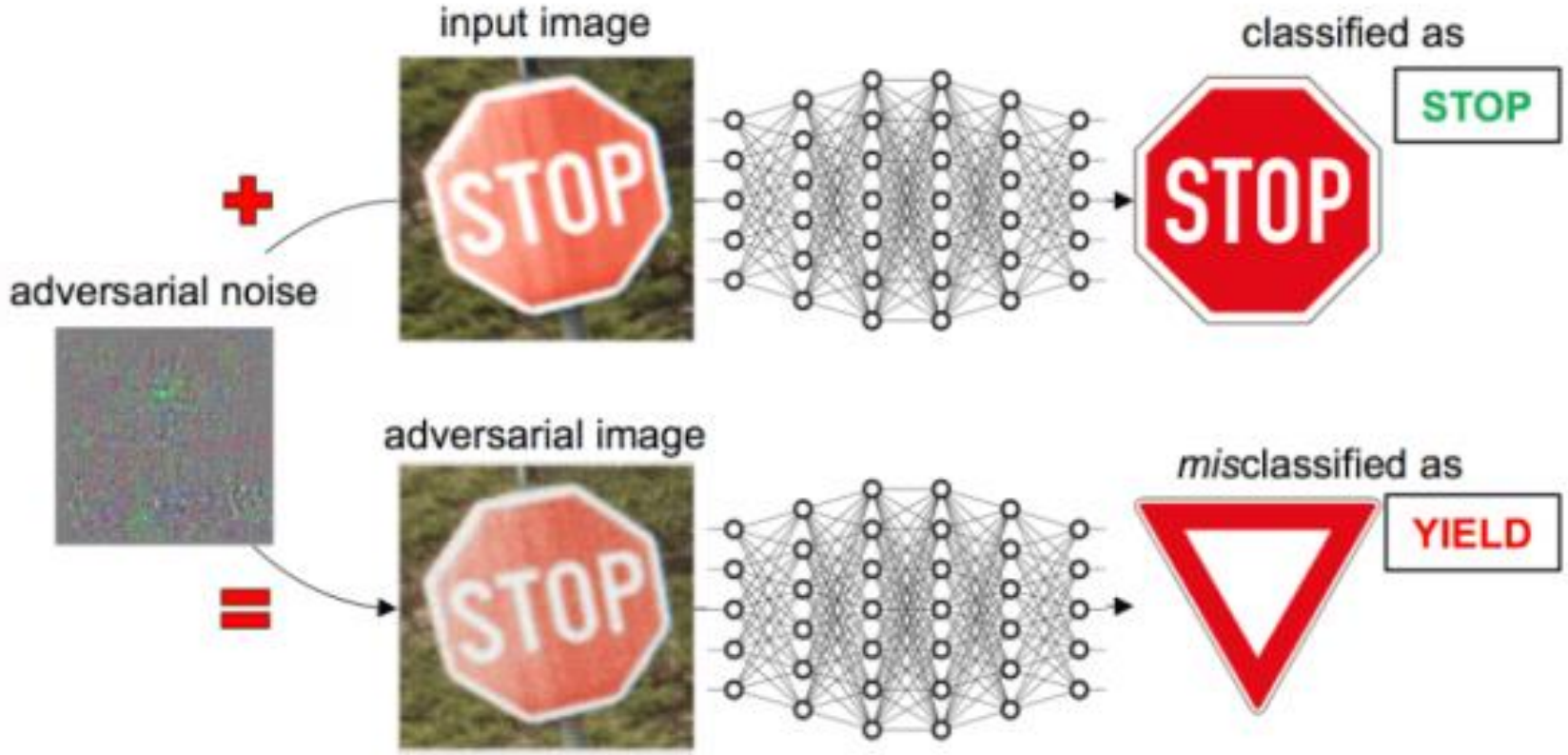
Introduction – Deep Learning

- DNNs for computer vision
  - > Convolutional Neural Networks: CNN
  - > Human-like accuracy



http://www.videantis.com/deep-learning-in-five-and-a-half-minutes.html

Introduction – Deep Learning



sondoesnotexist.com/
st-adversarial-attacks/

Introduction – Deep Learning

- Other challenges:
  - Blackbox nature of DNNs: high accuracy, but sometimes fail for no known reason
  - DNNs tune 100M+ parameters → not debuggable like human readable source code
- Conclusion
  - Gap between conventional V&V safety standards and nature of ML based systems
  - ISO 26262 is not applicable, different V&V methods are needed

Research Questions

1. What is the state-of-the-art in V&V of ML based safety-critical systems?
2. What are the main challenges when engineering safety-critical systems with DNN components in the automotive domain?

State-of-the-art in V&V for ML based SCS

- Incremental updates are not enough: new standards are needed
- Lots of research on innovative solutions and pioneering applications
  - > Standards lag behind
- Less research on engineering safety for DNNs, but growing interest
  - > V&V must be an integral part rather than an add-on

State-of-the-art in V&V for ML based SCS

- Open questions
  - How should a DNN component be classified? (SW unit or component)
  - Should DNN misclassifications be treated as HW failures? (Are ISO HW failure target values applicable?)
  - What metrics should be used to specify DNN accuracy? (Requirements specification target values?)

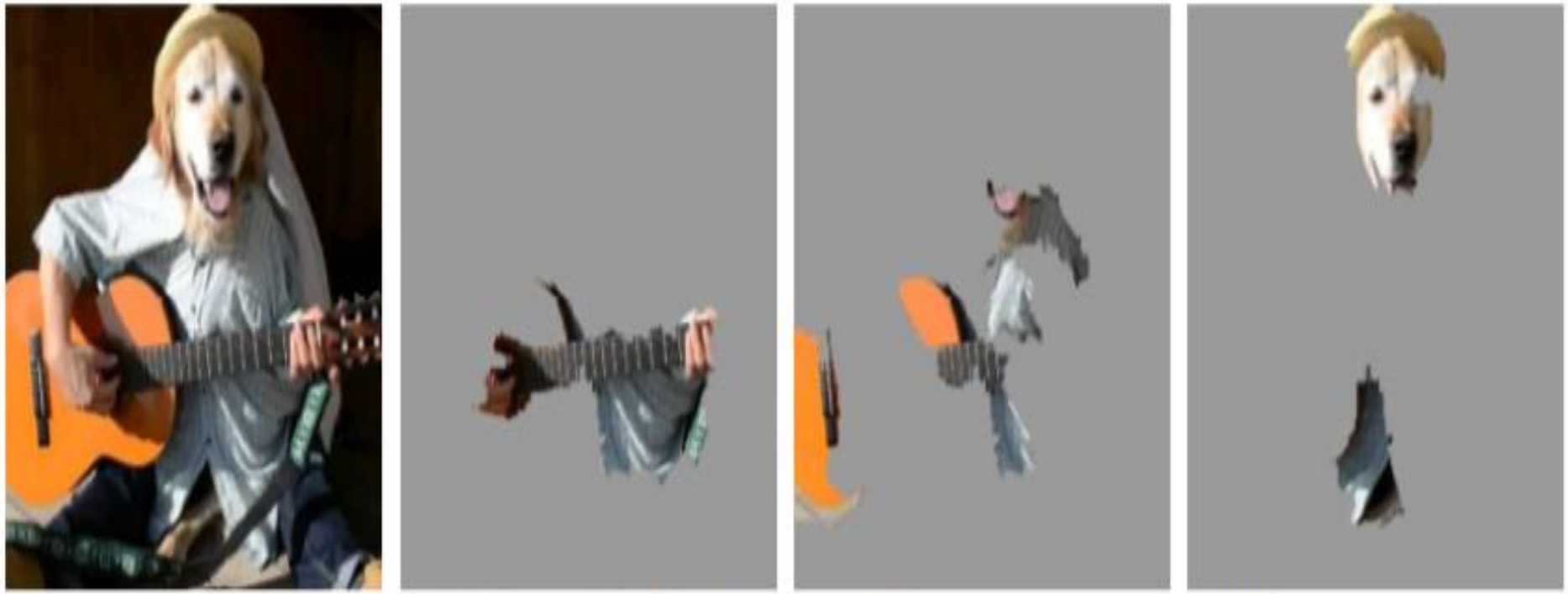
State-of-the-art in V&V for ML based SCS

- Academia and industry: agreement on the challenges, no agreement on best approach
- Academic approaches:
  - Formal methods (mathematic proofs)
  - Control theory (verification of learning behavior)
  - Probabilistic methods (statistical approaches)
- Industry approaches:
  - Simulated test cases→ test case design
  - Must cover normal and rare situations
  - Must protect against adversarial examples

State-of-the-art in V&V for ML based SCS

- Other possible V&V techniques and proposals:
  - Safety cage architecture
    - Continuous monitoring of inputs and uncertainties
    - Novelty detection
    - Redirecting execution to a “safe-track” if needed
  - Process guidelines
    - Training data collection
    - Testing strategies

State-of-the-art in V&V for ML based SCS



(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)

Main challenges of engineering SCSs with DNN components

- Several manufacturers actively engineer autonomous vehicles (Tesla, Uber, Volvo, etc.)
- But industry practice is far from being able to certify DNNs in SCSs
- Currently, safety-case augmentation requires human-in-the-loop

Main challenges of engineering SCSs with DNN components

- State space explosion
  - Varying environments
  - Varying quality of inputs
  - Ensuring correct behavior in all situations is difficult
- Importance of datasets
  - Geographical locations, city, country

Main challenges of engineering SCSs with DNN components

- Achieving robustness:
  - Goal
    - "Something you can trust" (FPs, FNs)
    - Requirements for training data and architecture should ensure robustness
  - Difficulty
    - Unpredictable environments
    - Robustness may require new definition in the context of DNN based autonomous vehicles

https://arxiv.org/pdf/1602.04938.pdf

Thank you for your attention!