Safety analysis and verification/validation of MachIne LEarning-based systems

Cristofer Englund

Research manager Cooperative systems, RISE Viktoria

Adjunct senior lecturer, Halmstad University

RISE Research Institutes of Sweden

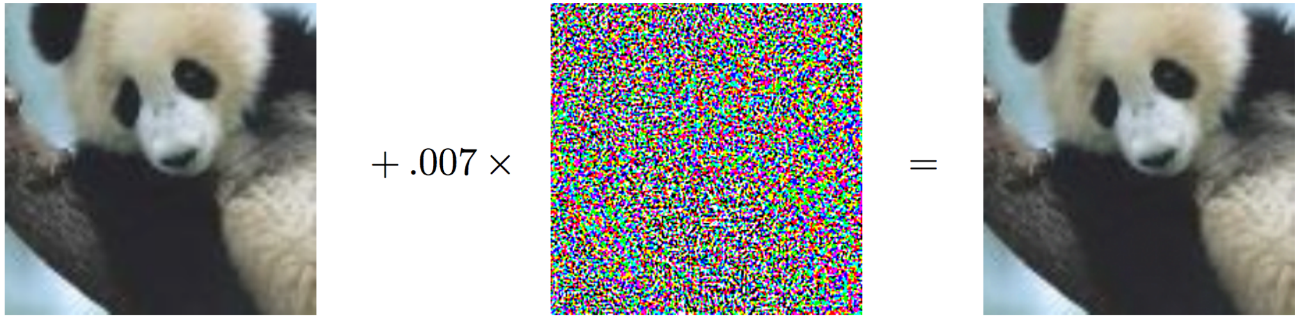
RISE ICT, Viktoria

Machine Learning in vehicles

- Why machine learning is necessary to enable autonomous driving
  - Traditional, rule-based, methods are static
  - Neural networks have the ability to generalize
- Trends that make Machine Learning possible in vehicles
  - Deep learning improves performance compared to traditional neural networks
  - Computational power for training and executing deep learning networks

Machine Learning – Neural Networks

- Neural Networks learns the desired behaviour from historical data
- We want the networks to generalize
  - The network should be able to take decisions on previously unknown data – if it is similar to the training data
- How do we avoid taking decisions on data that is not similar to training data?

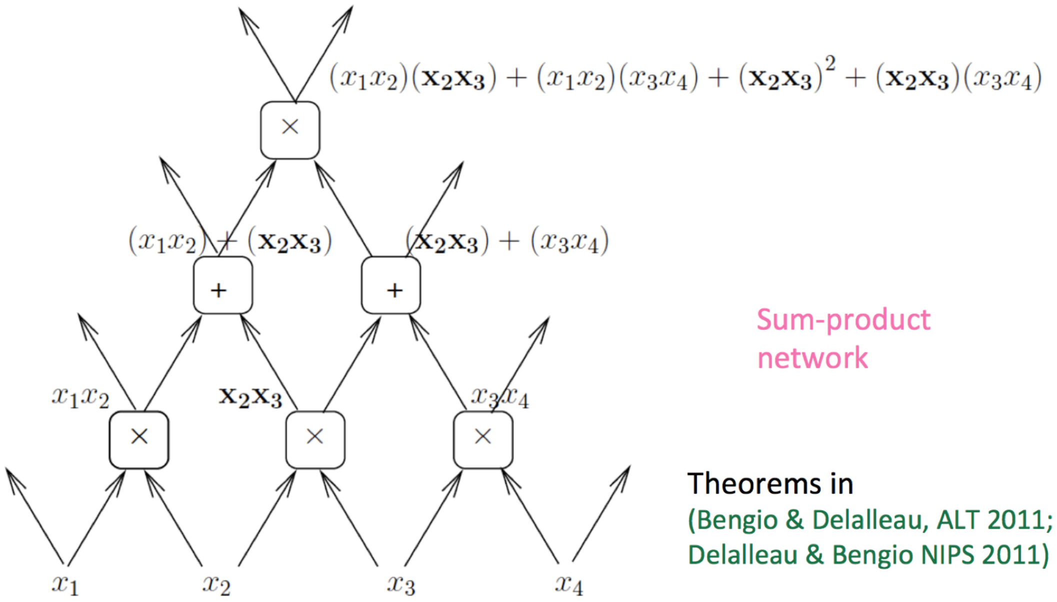


$$\boldsymbol{x} \quad + .007 \times \quad \mathrm{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y)) \quad = \quad \boldsymbol{x} + \epsilon \mathrm{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$$

$\boldsymbol{x}$	$\mathrm{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$	$\boldsymbol{x} +$ $\epsilon\mathrm{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"panda" 57.7% confidence	"nematode" 8.2% confidence	"gibbon" 99.3 % confidence

Why deep technologies?

- Polynomial expressed with shared components: advantage of depth may grow exponentially



Sum-product network

Theorems in
(Bengio & Delalleau, ALT 2011; Delalleau & Bengio NIPS 2011)

http://www.iro.umontreal.ca/~bengioy/cifar/NCAP2014-summerschool/slides/Yoshua_Bengio_CIFAR_school_12Aug2014.pdf

- Deep structures can make context mapping
- Deep Learning and deep knowledge

Existing knowledge

United Airlines' shares fell 8 percent yesterday, but rebounded by mid-day today.

United Airlines suffered from bad publicity due to mistreatment of passengers.

Ford' shares lost 3% because of the bad publicity caused by recent recalls.

New knowledge

United Airlines' shares fell 8 percent (possibly) because of the bad publicity

Neural Networks for perception

- To define the role of the NN it is important to have clear understanding about:
  - What role should the driver have?
  - What role should the system have?
  - Operational Design Domain ODD – the specific situations where a system is designed to operate in, e.g. a motorway or a geographical area.
- Local perception and awareness is key for AD
- Training NN to recognize hazardous situations
- Training NN to anticipate unforeseen situations

ML impacts on ISO 26262

- Five areas
  - Identifying hazards
  - Faults and failure modes
  - The use of training set
  - Level of ML usage
  - Required software techniques
- Hazard: ”a potential source of harm caused by malfunctioning behaviour of the item where harm is physically injury or damage to the health of persons”

An Analysis of ISO 26262: Using Machine Learning Safely in Automotive Software. Rick Saly, Rodrigo Queiroz, Kryzsztof Czarnecki. arXiv: 1709.02435v1

ML impacts on ISO 26262

- Identifying hazards
  - Automation takes over more and more control.
  - Taking over becomes increasingly critical.
  - Increased automation can/will create behaviour change in the operator –> reducing their skill level.

  → Include harm potentially caused by complex behaviour interaction between human and vehicle.
- Faults and failure modes
  - Incorrect output for a given input.

  → Current recommendations apply.
- The use of training set
  - Necessary to use ML for perception. A training set is used instead of a specification.
  - Data does not contain all possible scenarios.

  → Design systems that can cope with an error rate.
- Level of ML usage
  - End-to-End systems model all functionality and the result is a complex black box system.

  → Use ML at the component level.
- Required software techniques
  - ISO 26262 requires many specific techniques for software development.
  - Some apply to ML, some may be adapted but some others assume programming.

  → Express requirement in terms of intent and maturity of the techniques rather than their specific details.

An Analysis of ISO 26262: Using Machine Learning Safely in Automotive Software. Rick Saly, Rodrigo Queiroz, Kryzsztof Czarnecki. arXiv: 1709.02435v1

What is SMILE II about?

- We accept that DNN are black-boxes and that we need to include them in vehicle perception
- Camera-based perception models
- Investigate pre-training
  - Humans learn from birth what is “dangerous”
  - Can self driving vehicles make use of other contexts?
- Investigate how to handle model updates
- Demonstrate perception use-case

RI.
SE

What is SMILE II about?

- Safety Cage to monitor the data presented to the network

SMILE focus initially on
Input data analysis

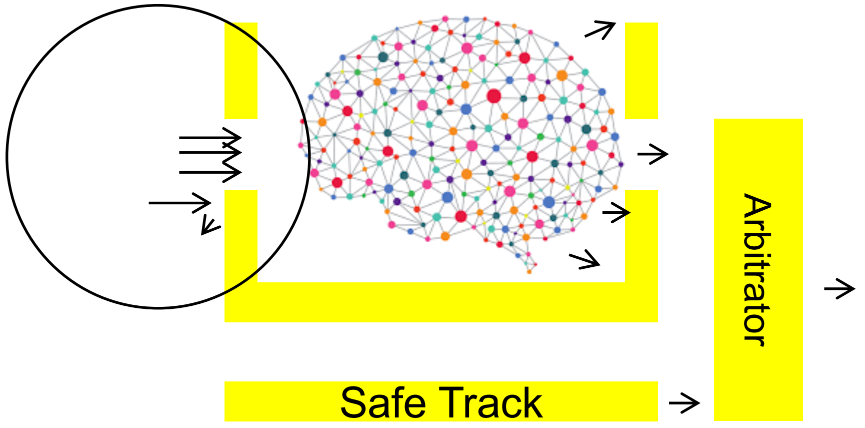
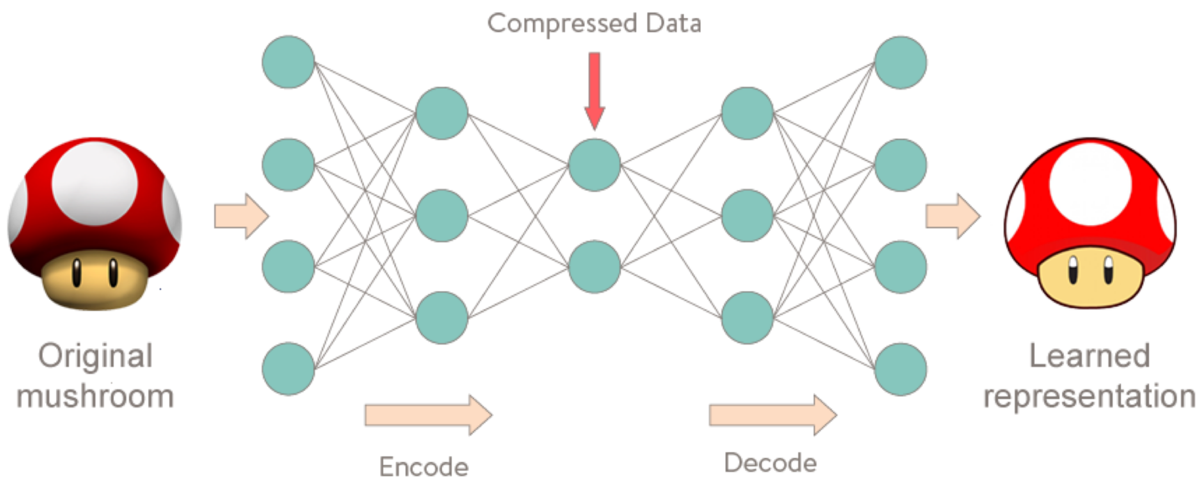


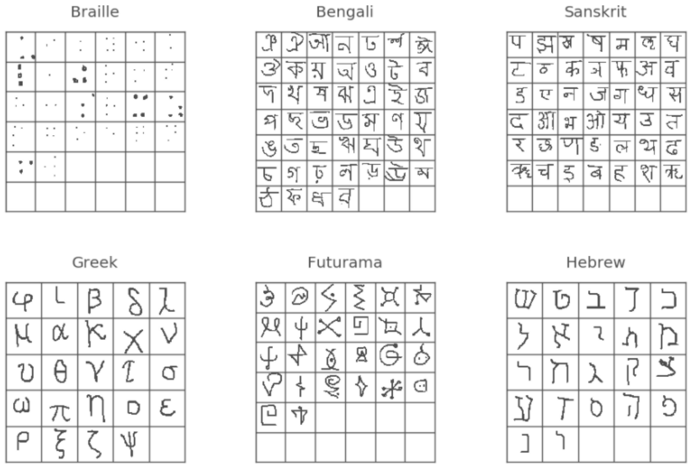
Image anomaly detection using convolutional autoencoders



Work by: Lars Tornberg, VCC

Data set

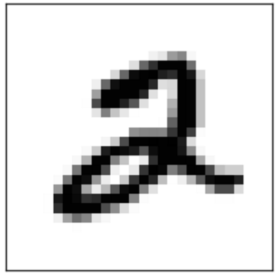
MNIST data set:
- 70 000 images (28x28)



Omniglot data set:
- 1623 images (105x105)
- 50 different alphabets
- 20 examples per char

Data preparation

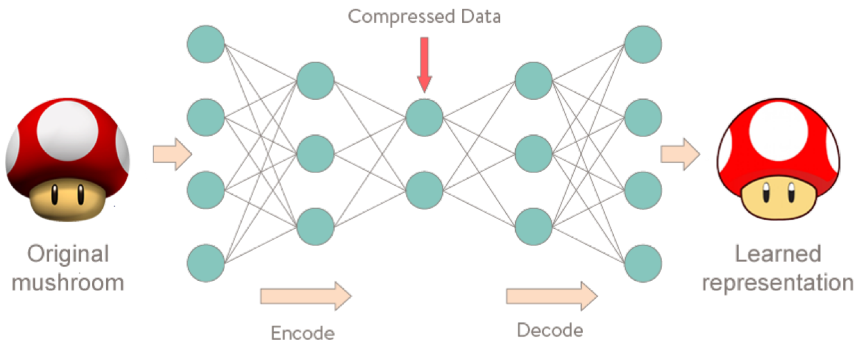
Resize omniglot images to MNIST size (105x105 -> 28x28):

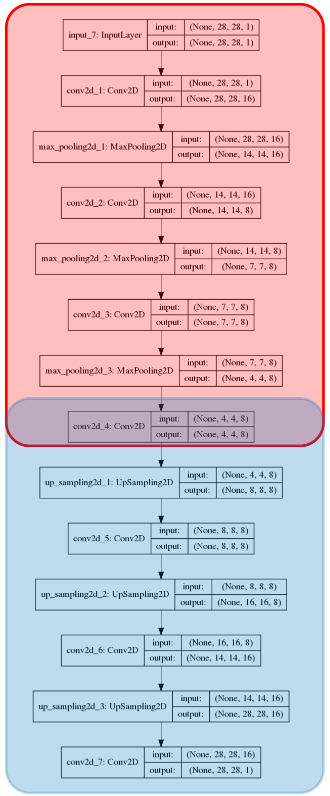


Nearest neighbour interpolation

Bilinear interpolation

Convolutional Autoencoder

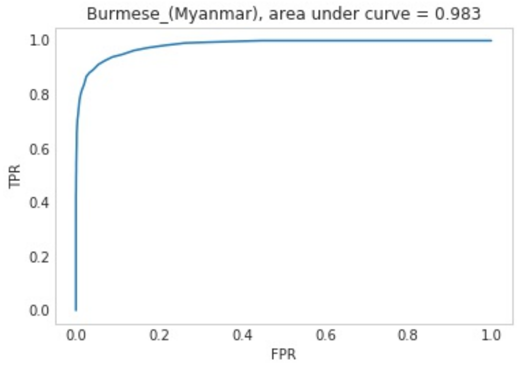


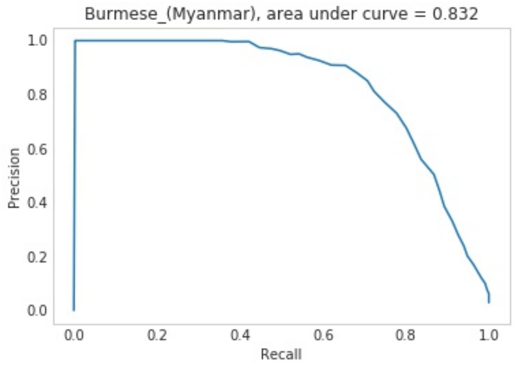


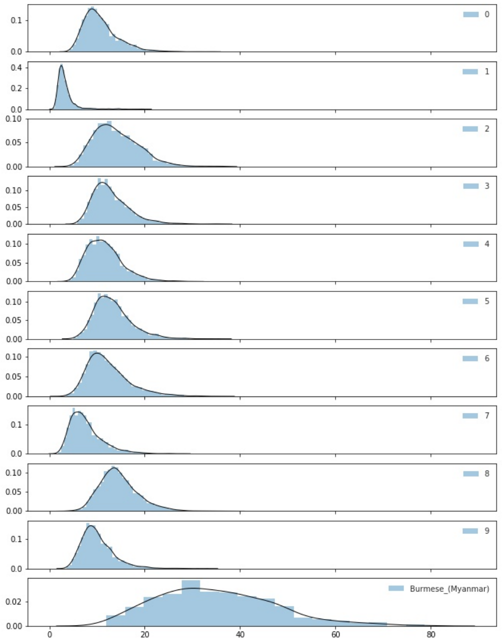
Framework: Keras
Loss: Pixel by pixel MSE
Training set: MNIST 46900 examples (67%)

Method evaluation

Burmese (Myanmar), NN-interpolation





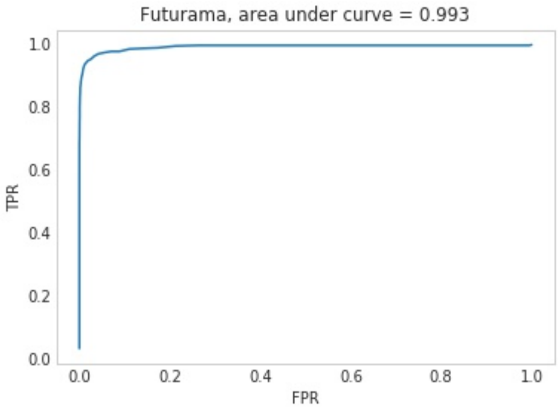


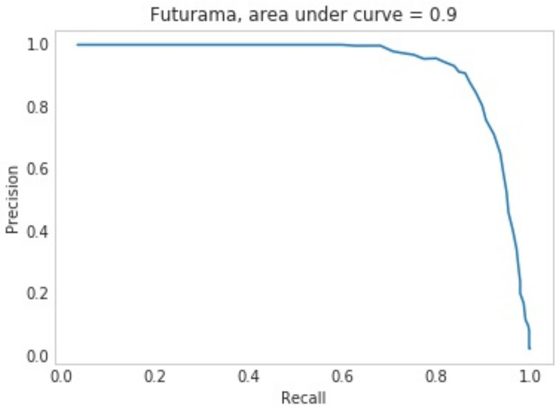
$$\text{Precision} = \frac{tp}{tp + fp}$$

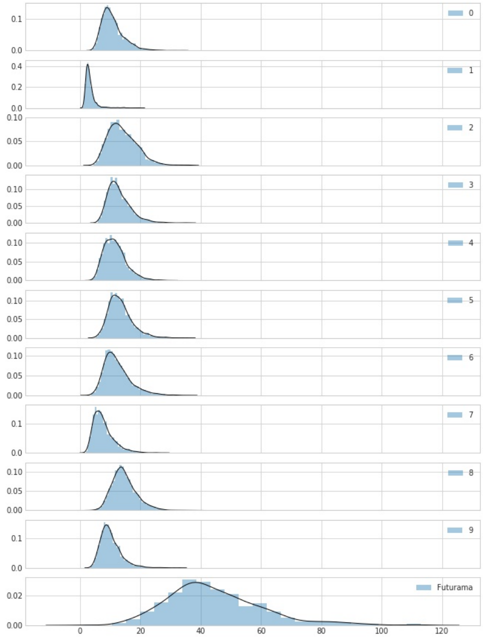
$$\text{Recall} = \frac{tp}{tp + fn}$$

Method evaluation

Futurama, NN-interpolation





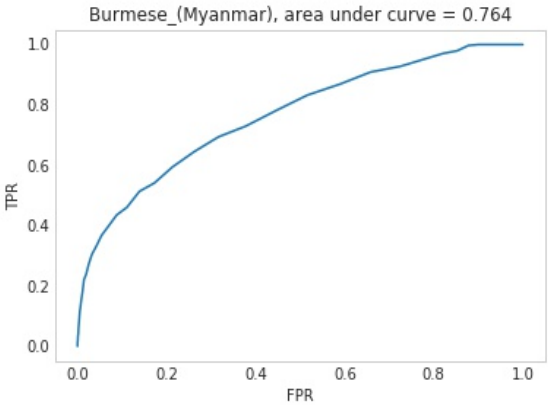


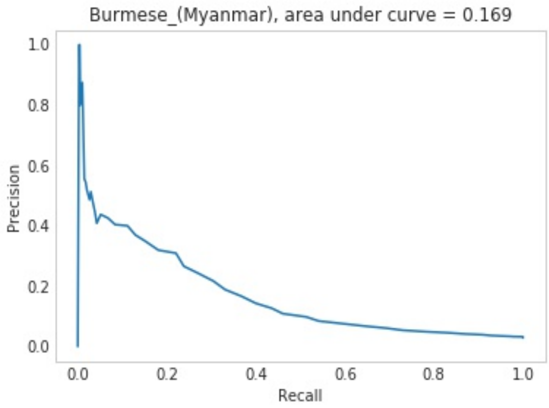
$$\text{Precision} = \frac{tp}{tp + fp}$$

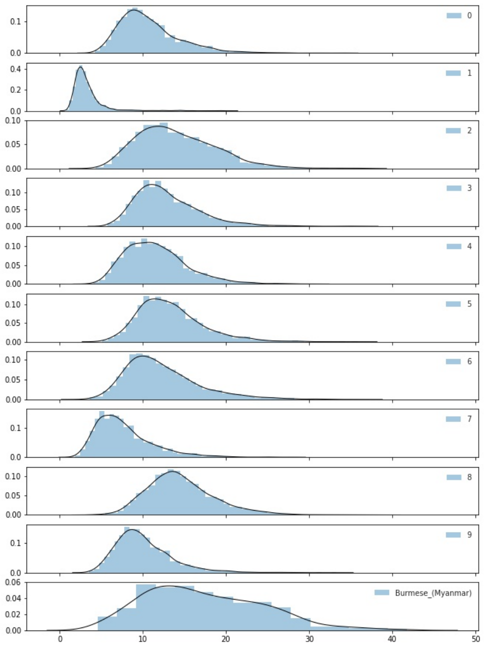
$$\text{Recall} = \frac{tp}{tp + fn}$$

Method evaluation

Burmese (Myanmar), bilinear-interpolation





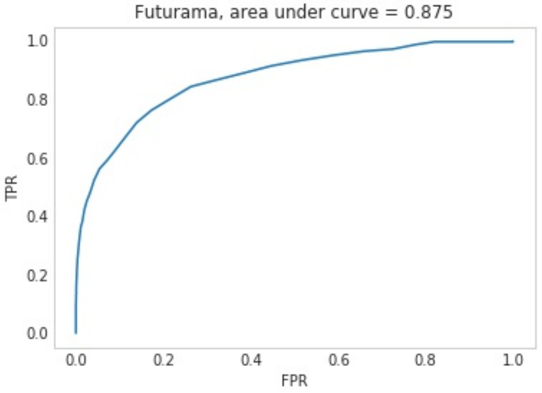


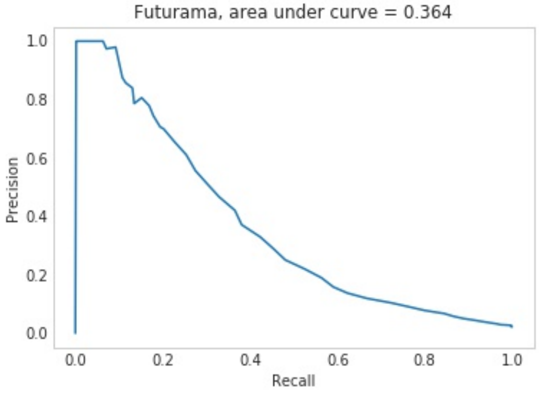
$$\text{Precision} = \frac{tp}{tp + fp}$$

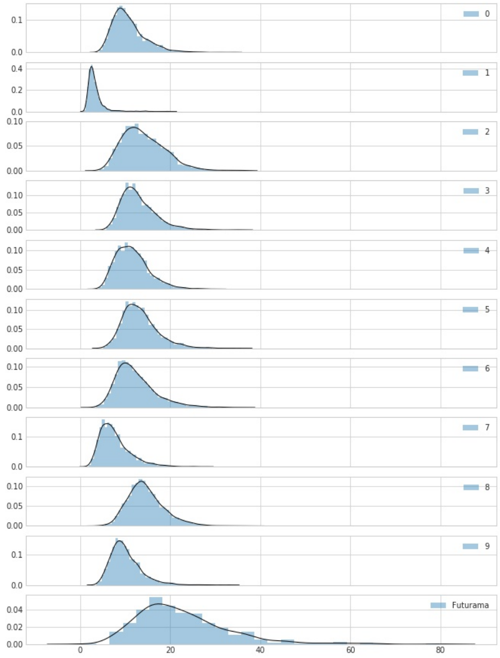
$$\text{Recall} = \frac{tp}{tp + fn}$$

Method evaluation

Futurama, bilinear-interpolation







$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

Results and conclusion so far

- It is important to understand the input space
  - What data should the network be allowed to process?
  - How should the data be pre-processed?
- Understand what is unique in the images
  - How different from the training data is ok?
- Developing the safety-cage using “simple” datasets can prove soundness of the method, but must also be thoroughly evaluated in the final domain.
  - How is difference estimated in high dimensional space?
- Publications

  - Henriksson, J., Borg, M., Englund, C.: **Automotive safety and machine learning: Initial results from a study on how to adapt the ISO 26262 safety standard**. In: SEFAIAS-2018. (2018)
  - Borg, M., Englund, C., Duran, B.: **Traceability and Deep Learning - Safety-critical Systems with Traces Ending in Deep Neural Networks**. In: In Proc. of the Grand Challenges of Traceability: The Next Ten Years. (2017) 48–49
  - Englund, Cristofer; Borg, Markus; Duran, Boris; Kaijser, Henrik ; Lönn, Henrik; Lindström, Konstantin; Zandén, Carl; Levandowski, Christoffer; Simoen, Michaël; Törnquist, Jonas. **Deep Learning and Safety-critical Systems: Research, Practice, and Future Needs in Automotive.** In review IEEE Transactions on Intelligent Transportation Systems

Safety analysis and verification/validation of MachIne LEarning- based systems

Cristofer Englund

cristofer.englund@ri.se

RISE Research Institutes of Sweden

RISE ICT, Viktoria