EXPERIENCES FROM APPLYING CONVOLUTIONAL NEURAL NETWORKS FOR CLASSIFYING 2D SENSOR DATA

ROSEN
empowered by technology

ABOUT ME

Matthias Peußner

Applied System-Science

mpeussner@rosen-group.com

Software Developer/ Data Scientist

Slides: https://github.com/rosen-group/conferences

INTRODUCING THE ROSEN GROUP

ROSEN develops and manufactures equipment, software, and methods for the inspection, diagnosis, and protection of industrial structures in a wide range of industries.

Because damage can cause serious impacts!

INSPECTION OF PIPELINES

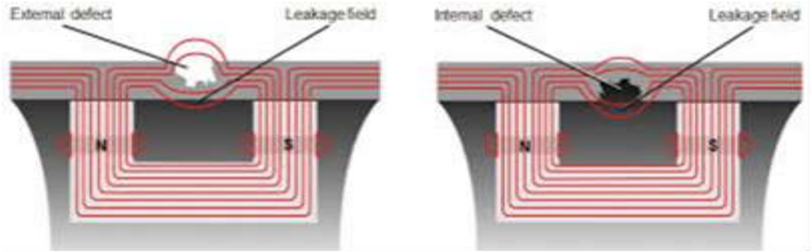
We inspect industrial assets…

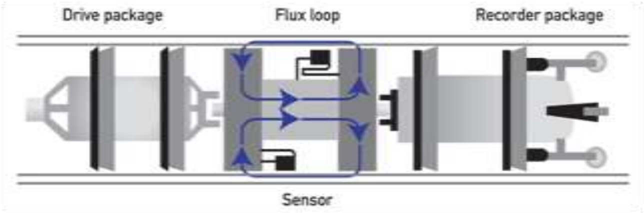
- **100.000 miles/a (160.000 km/a)**

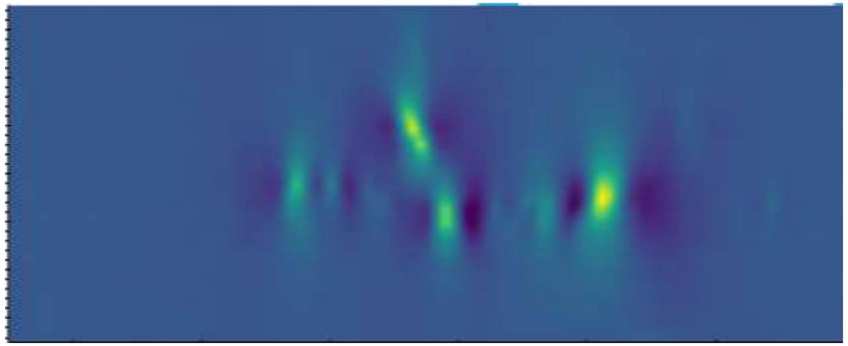
MAGNETIC FLUX LEAKAGE

- Measure volume loss in pipeline wall
- Indirect measurement principle
- Image-like data (2d array of amplitudes)

- Tasks: Detect, classify and estimate defect geometry from measured data.





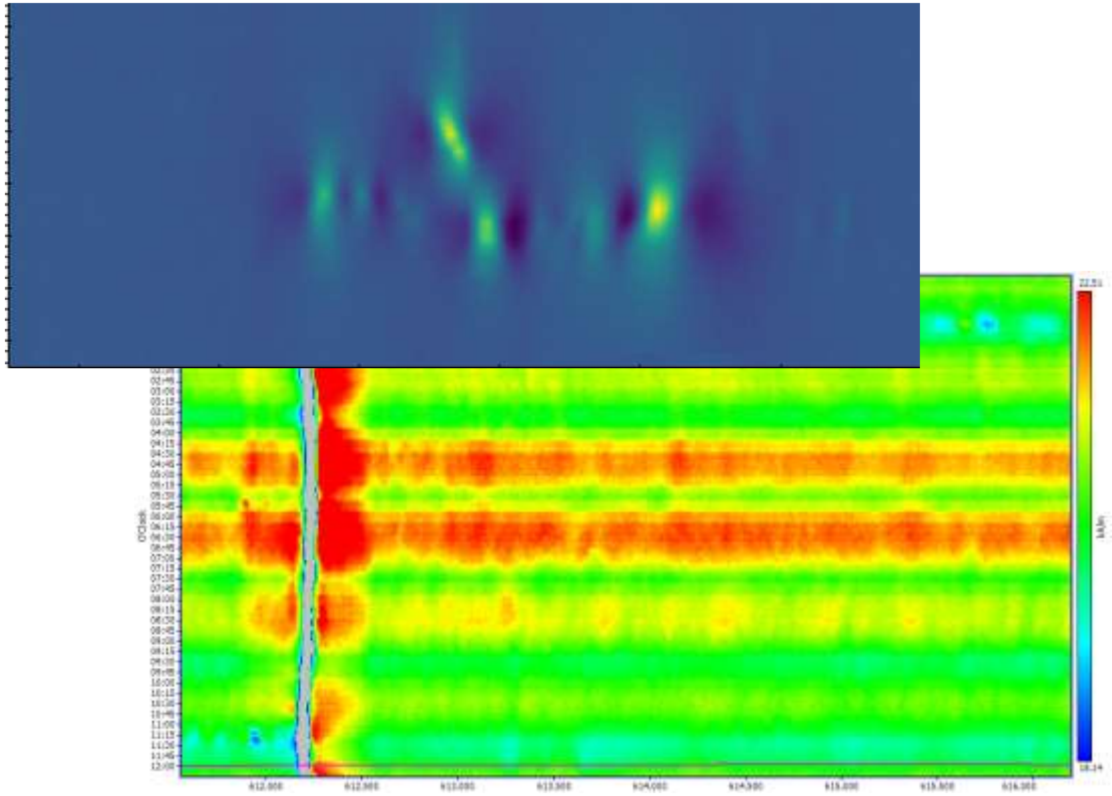


SENSOR SIGNAL AND IMAGES

Magnet Flux Leakage (MFL) technology

Channels: magnetic field value

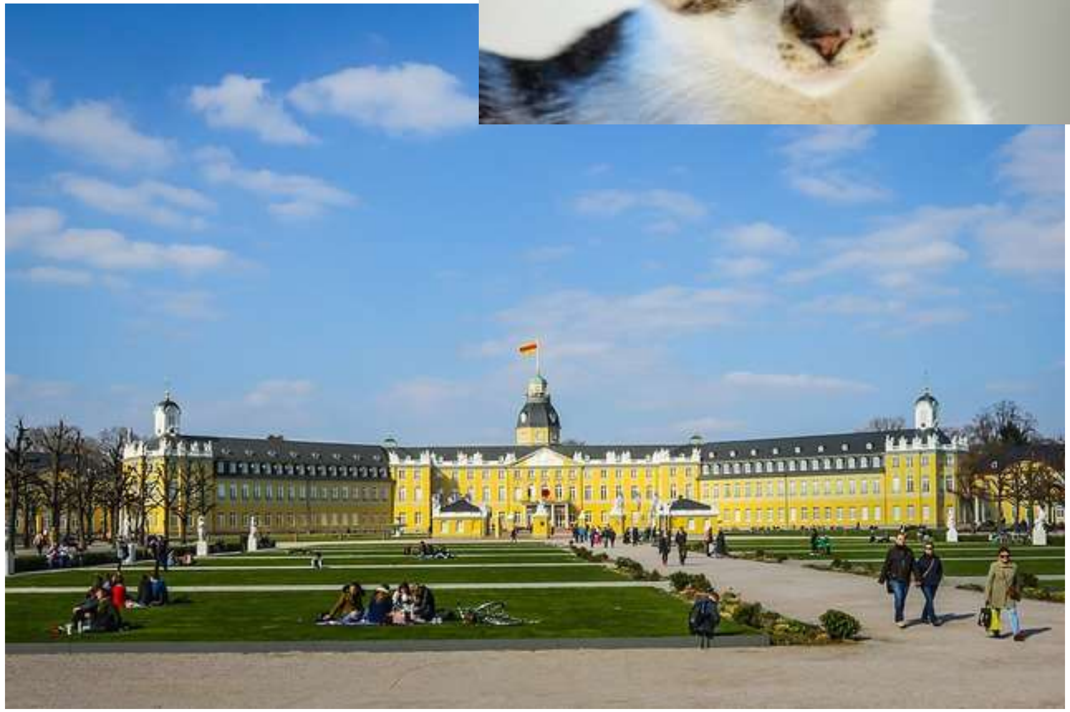
Value range: 3.500-50.000 (exemplary)



Images

Channels:
Red, Green, Blue

Value range: 0-255



PIPELINE DEFECTS

START

Investigation started in **2014** as conceptual study/ technology study

Can our evaluation algorithms benefit from using deep neural networks (DNNs)?

Do we have enough data?

START

Tutorial from Caffe

LeNet MNIST Tutorial

Image Classification

Convolutional Neural Network (CNN)

Deep Learning Library (with GPU support)

Caffe, http://caffe.berkeleyvision.org/, python interface

Hardware

Nvidia Quadro K4000

Now there are several other Deep Learning frameworks available, e.g. Keras, Tensorflow

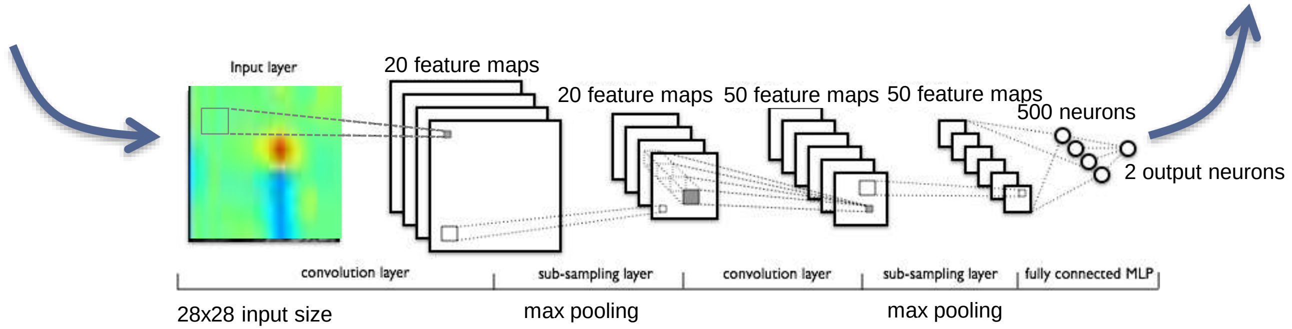
NETWORK ARCHITECTURE

Input Data

Patches

- Sensor Channel

Output
'defect' or 'no-defect'



Training

Drop-out was used during training.

Batch size of 100.

Trained from scratch. No pre-training.

Few hours for one training.

Number of labeled Test and Training Samples

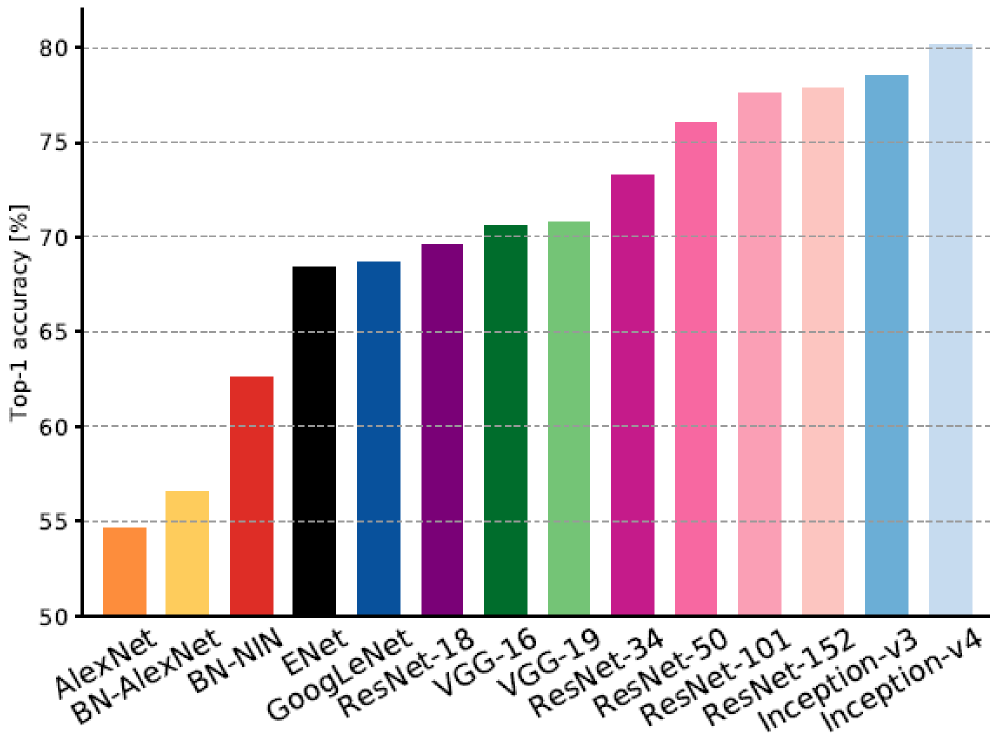
- 840.000 from 50 different pipelines

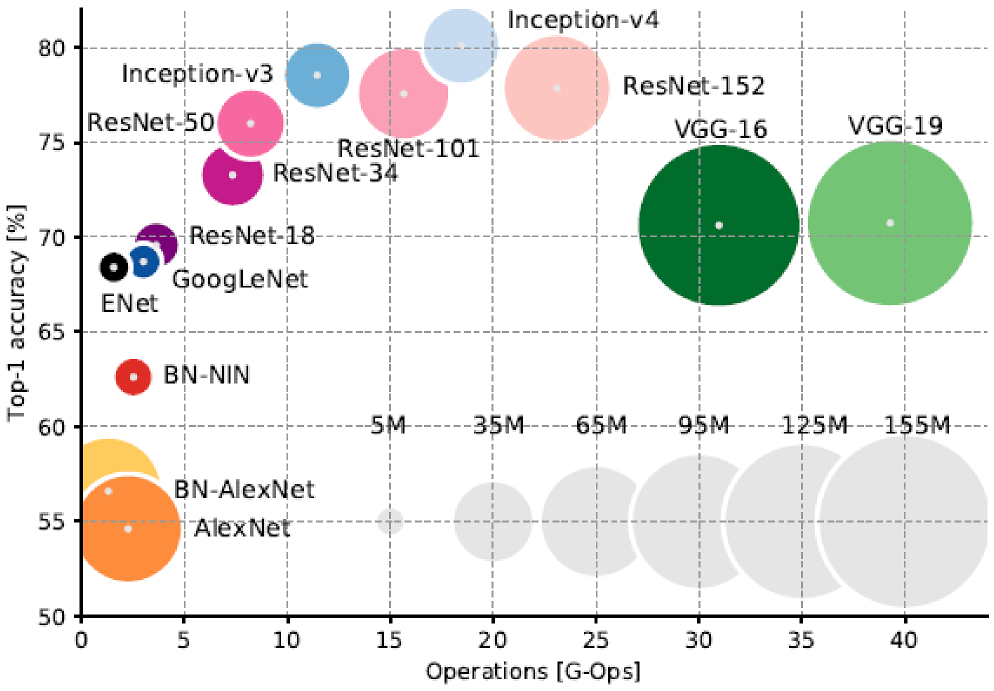
Samples were manually labeled by experts.

Architecture is similar to LeNet (1998)

Very small compared to current architectures

NETWORK ARCHITECTURE SELECTION





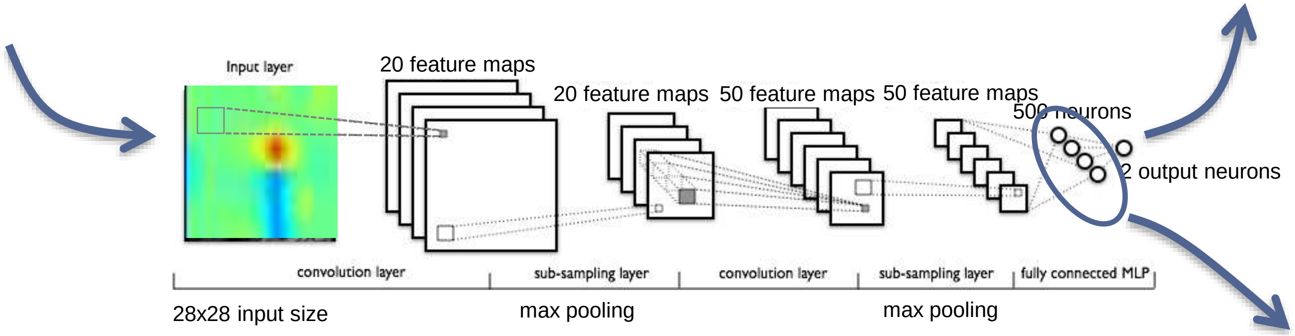
Source: A. Canziani, A. Paszke and E. Culurciello, “An Analysis of Deep Neural Network Models for Practical Applications”, *CoRR*, 2016

CNN AS FEATURE EXTRACTOR

Input Data

Output

‘Anomaly’ or ‘False Call’



Alternatively

UMAP,

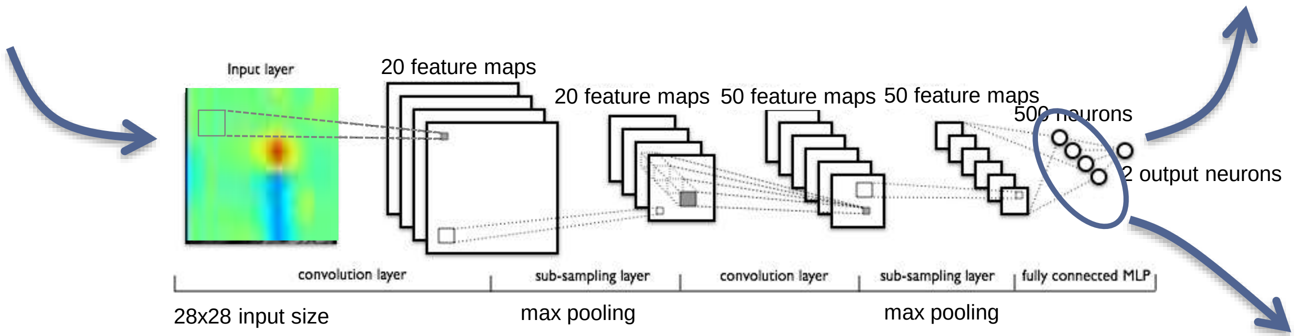
Uniform Manifold Approximation and Projection

CNN AS FEATURE EXTRACTOR

Input Data

Output

‘Anomaly’ or ‘False Call’



t-SNE

unsupervised

Dimension reduction from 500 to 2

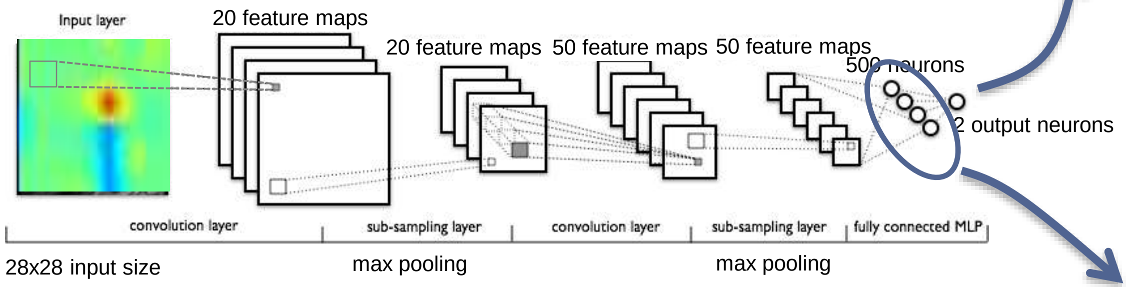
Alternatively

UMAP,
Uniform Manifold Approximation and Projection

CNN AS FEATURE EXTRACTOR

Input Data

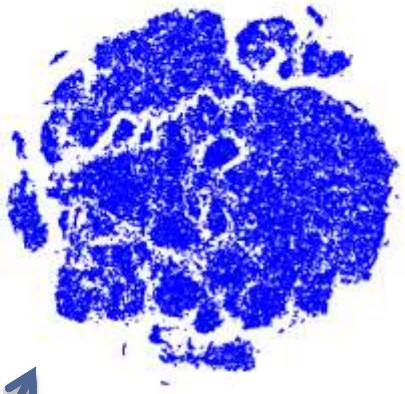
Output
'Anomaly' or 'False Call'



t-SNE

unsupervised

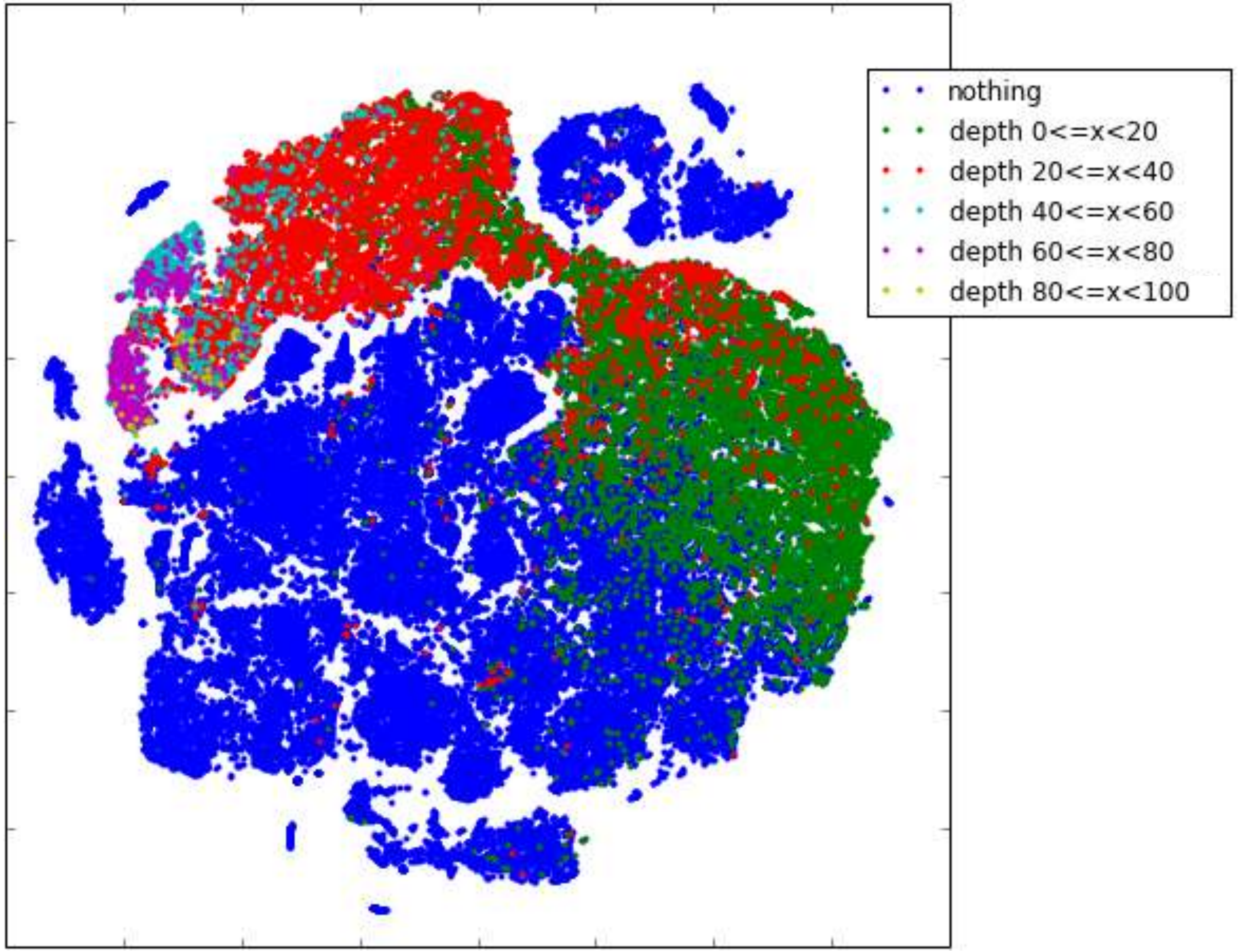
Dimension reduction from 500 to 2



Alternatively

UMAP,
Uniform Manifold Approximation and Projection

CNN CODE EMBEDDING

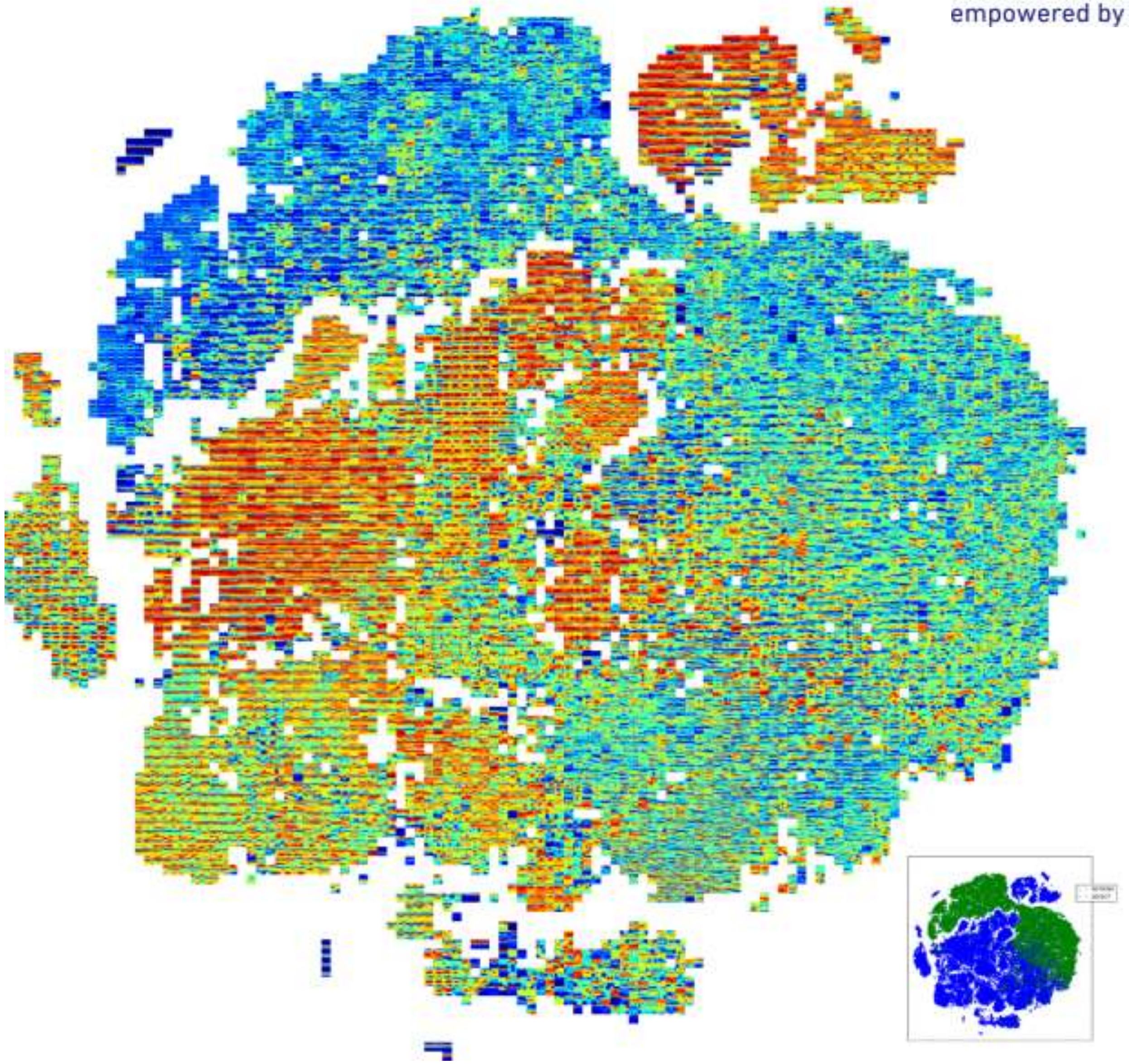


Color shows relative depths of defects

70,000 data points

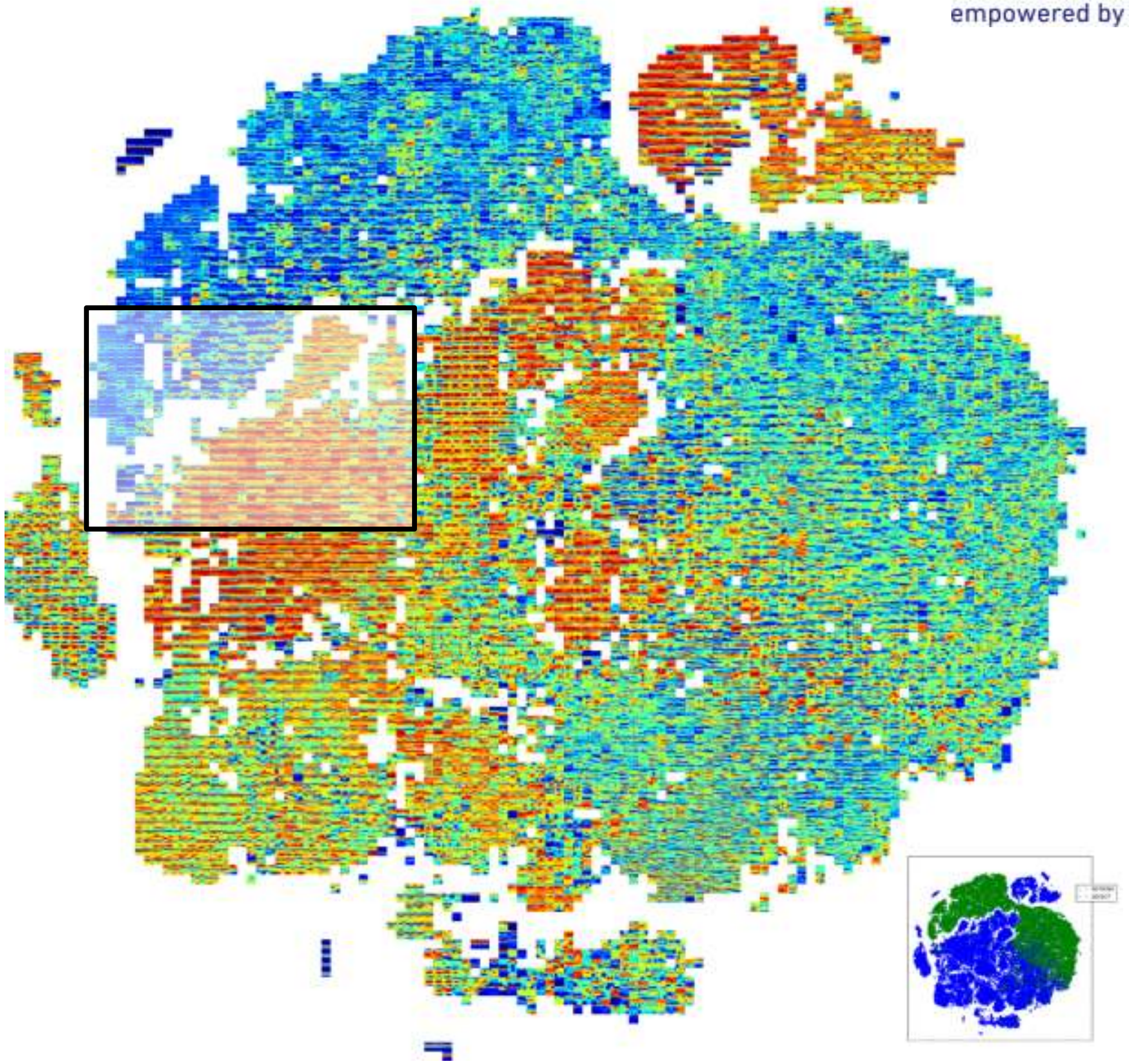
CNN CODE EMBEDDING

- Showing the sensor signal patches instead of just a dot within the t-SNE Embedding



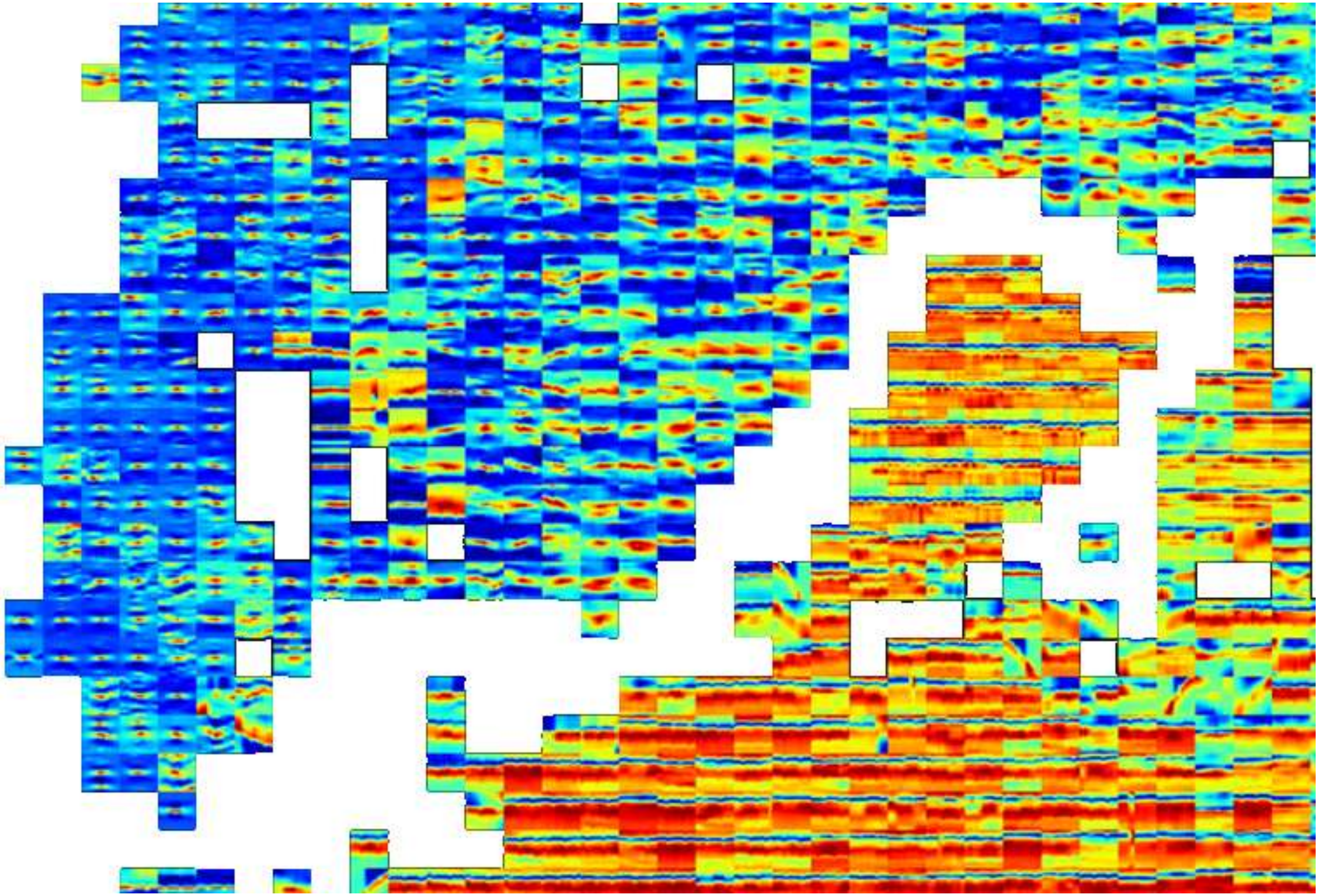
CNN CODE EMBEDDING

- Showing the sensor signal patches instead of just a dot within the t-SNE Embedding

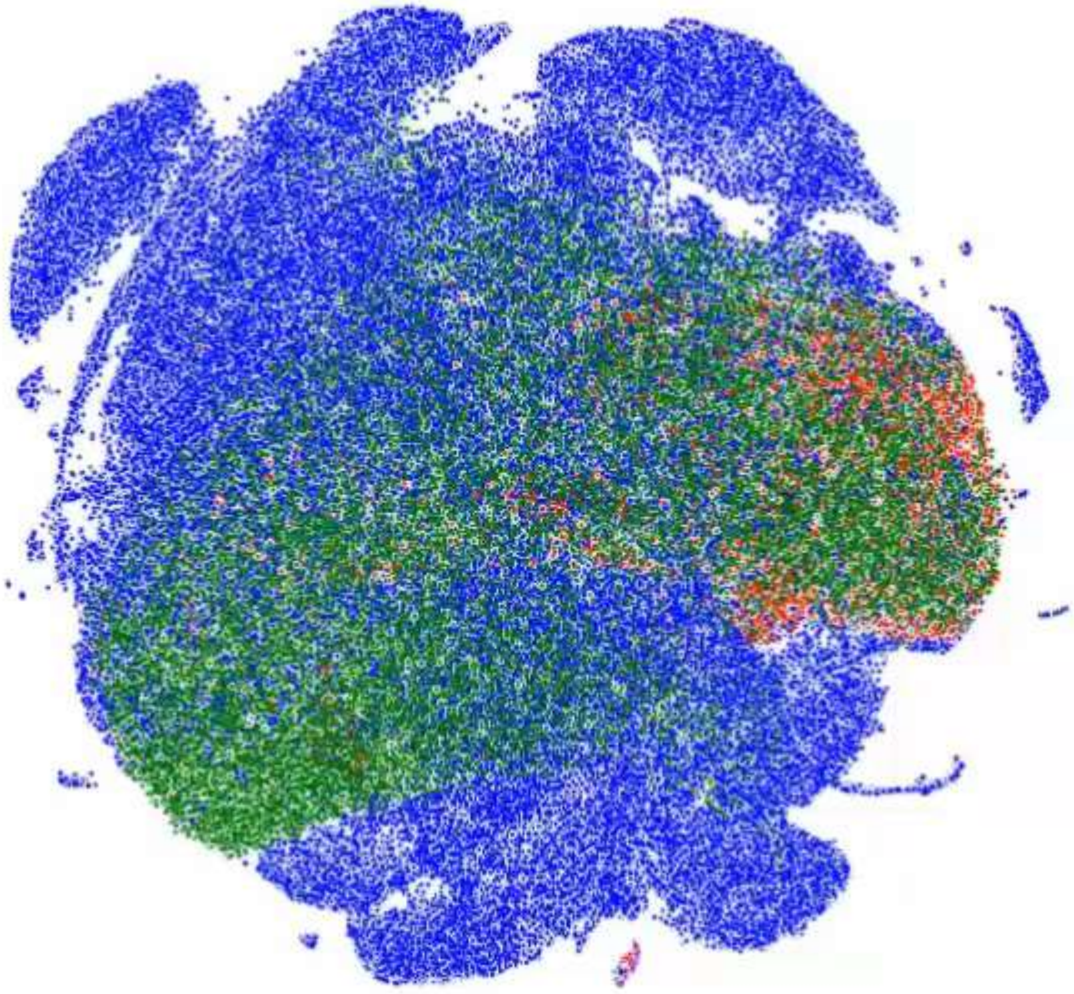


CNN CODE EMBEDDING

- Zoom



CNN CODE EMBEDDING



AUTOMATE YOUR PROCESSING

Because we haven’t done it only once:

- Pre-processed, converted and augmented the training and test data
- Trained and tested a network
- Analyzed the network and the test results

we automated the processing using

- Batch scrips
- Python scripts using the luigi module

- … to minimize errors in the workflow
- … to not repeat a processing step with the same parameters twice
- … to parallelize and distribute work
- … to foster repeatability
- … to monitor progress

Alternative: Apache Airflow

NETWORK INPUT

How large should be the input patch, in meter not pixel size?

- Information not present in the patch cannot be considered by the network

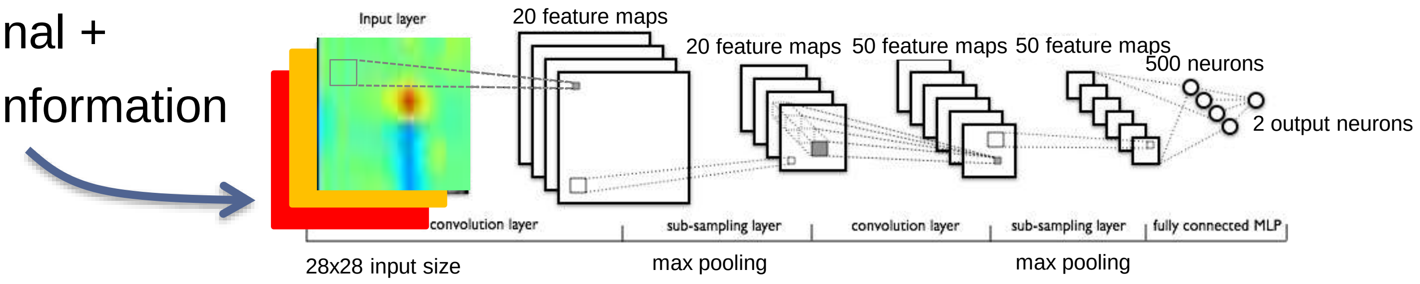
An experienced evaluator takes more information into account

- Surround of the defect
- History of the pipeline

Additional information can be added as further input layers or as further input to the fully connected layers.

Sensor Signal +

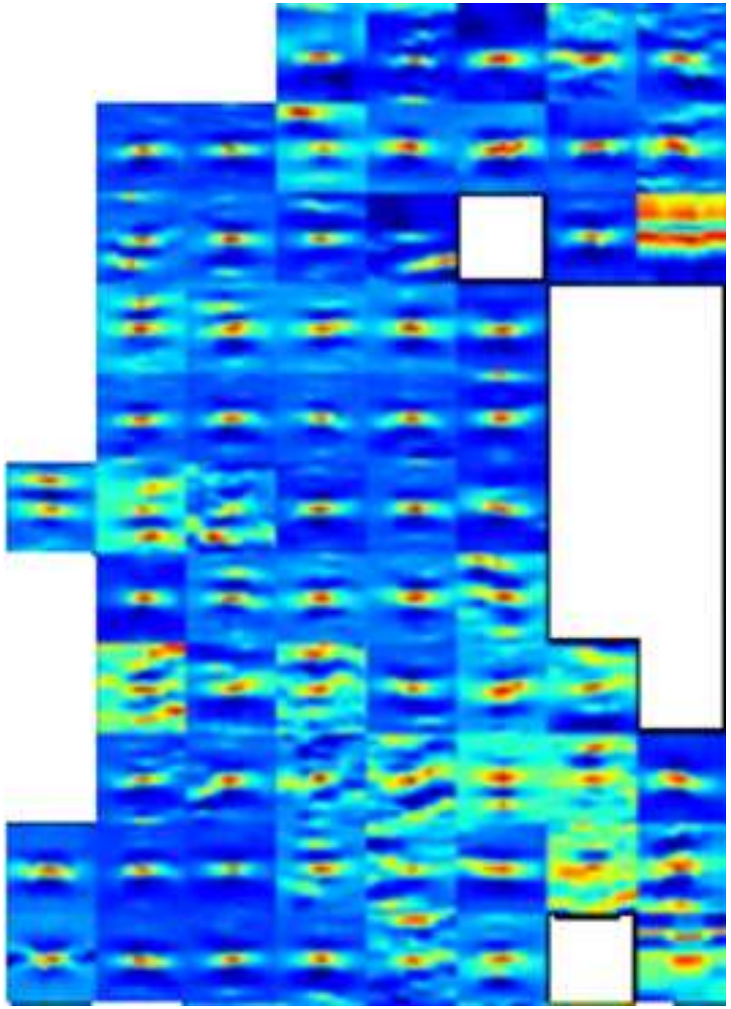
Additional information



TEST SET

Train-Test set split: Training 90%, Testing 10%

- Class Imbalance
  - Defect vs. no-defect
  - Shallow defects vs deep defects
  - Oversampling (duplicate deep defects)
- Shuffling



ACCURACY

We got an accuracy of around 93%.

Is 93% sufficient?
What if mostly deep defects are incorrectly classified?
Are there any wrongly labeled test data?
What is the human performance?

ACCURACY

We got an accuracy of around 93%.

Is 93% sufficient?
What if mostly deep defects are incorrectly classified?
Are there any wrongly labeled test data?
What is the human performance?

You have to define what you want to optimize!

7% of the test data are ~5.800 samples.
You may need the resources from experts to analyze them. Are there any systematic errors?

ACCURACY

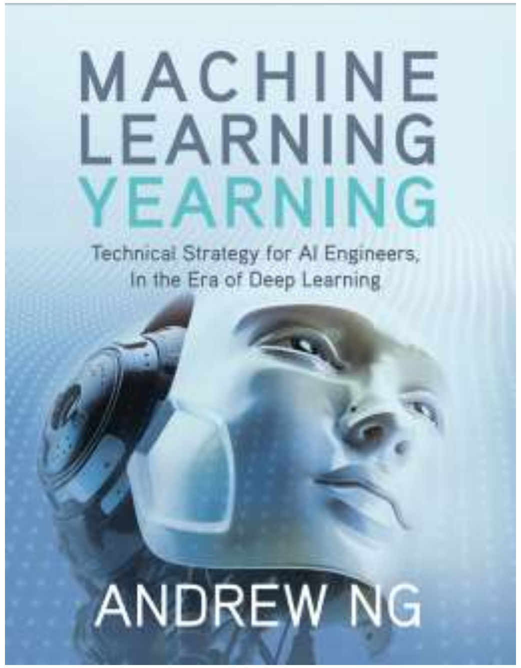
We got an accuracy of around 93%.

Is 93% sufficient?
What if mostly deep defects are incorrectly classified?
Are there any wrongly labeled test data?
What is the human performance?

You have to define what you want to optimize!

7% of the test data are ~5.800 samples.
You may need the resources from experts to analyze them. Are there any systematic errors?

Recommended Literature



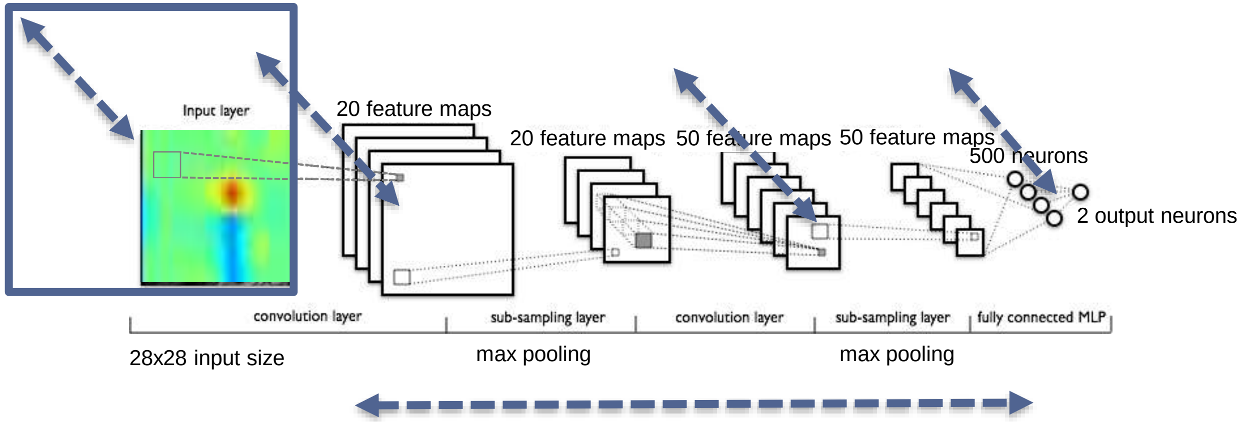
ACCURACY

Can we get better?

Tests in varying the network architecture

- adding or removing one convolutional layer
- increasing the input size e.g. 120x28 pixels

Changes the accuracy (slightly): 93% +/- 2%



Another option instead of modifying standard architectures manually: neural architecture search (NAS)

ACCURACY

Leave-one-'pipeline'-out cross validation :

Training samples come from 50 pipelines. Training on 49 pipelines, testing on 1 pipeline.

This is our typical use case:

Inspect an unseen pipeline using the knowledge from previous inspected pipelines.

ACCURACY

Leave-one-’pipeline’-out cross validation :

Training samples come from 50 pipelines. Training on 49 pipelines, testing on 1 pipeline.

This is our typical use case:

Inspect an unseen pipeline using the knowledge from previous inspected pipelines.

Accuracy varies between 35% and 99%, depending on the testing pipeline!

ACCURACY

Leave-one-’pipeline’-out cross validation :

Training samples come from 50 pipelines. Training on 49 pipelines, testing on 1 pipeline.

This is our typical use case:

Inspect an unseen pipeline using the knowledge from previous inspected pipelines.

Accuracy varies between 35% and 99%, depending on the testing pipeline!

Can we generalize from one pipeline to the other?

Actually we have hidden parameters, depending on the pipeline or the inspection itself.
Reminder: Our training set contains 840.000 patches from only 50 pipelines.

FIRST FINDINGS

Investigation started in 2014 as conceptual study/ technology study

Can our evaluation algorithms benefit from using deep neural networks (DNN)?

A feature extractor can be build much faster than it is handcrafted.

The expert currently has more information on a patch than the network.

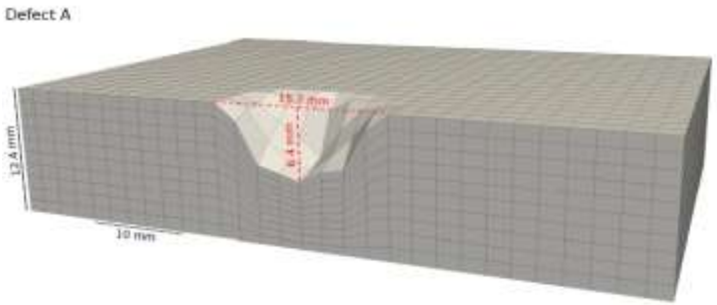
How to generalize over pipelines?

Do we have enough data?

Yes. But we do not get them for free. We have to invest in building up a training and test set.

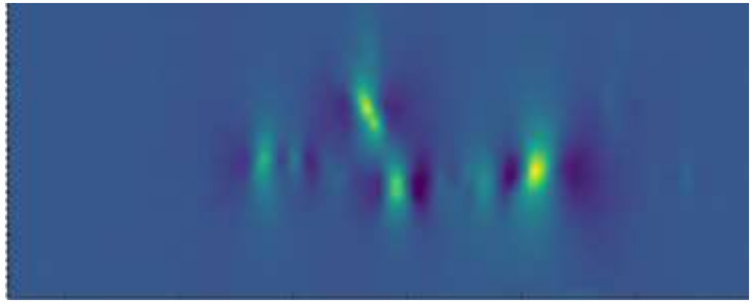
ONGOING INVESTIGATIONS

- Take more data from different pipelines
- Transfer Learning
- Create synthetic data
  (see Talk Synthetic data for machine learning, Hendrik Niemeyer, PyCon.DE 2017)



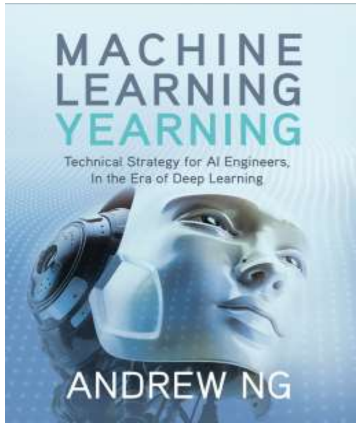
TAKE AWAYS

You can get your first results fast. It is easy to start.



Signal data may be less intuitive for human perception but less complex for a neural network than natural photos scenes with animals and people.

If you cannot rely on human perception, you have to think about how to get your ground truth data.



Creating your dev and test sets is crucial to your projects success.

Maybe you need a high effort to change and improve your dev/test sets within your project.

Working on your training and test data sets may take more time than actually the training.

THANK YOU FOR JOINING THIS PRESENTATION.

ROSEN
empowered by technology