CQM
Consultants in Quantitative Methods

inspectation

Safe and efficient inspection of railway tracks using deep learning models

Dr. Johan van Rooij, 16-01-2019

The Speaker: dr. Johan van Rooij

- Until December: Senior Consultant

CQM
Consultants in Quantitative Methods

- Just started as: Senior Data Scientist

VALID
STAY AHEAD

- 1 day a week: Assistant Professor

Universiteit Utrecht

- PhD (of the LNMB) in Theoretical Computer Science (Algorithms).
- Very broad interest in applications of mathematics.
- Trying to be *“a guide in the ever changing world of applied mathematics and data science”*.
- Reach me at: **J.M.M.vanRooij@uu.nl**

Hendrik Lorentz Prize – Nederlandse Data Science Prijzen

- “De Hendrik Lorentz Prijs is bedoeld voor een organisatie binnen het bedrijfsleven of de overheid die op een onderscheidende en innovatieve manier data science toepast.“

KONINKLIJKE HOLLANDSCHE MAATSCHAPPIJ DER WETENSCHAPPEN

Big Data Alliance

NEDERLANDSE DATA SCIENCE PRIJZEN

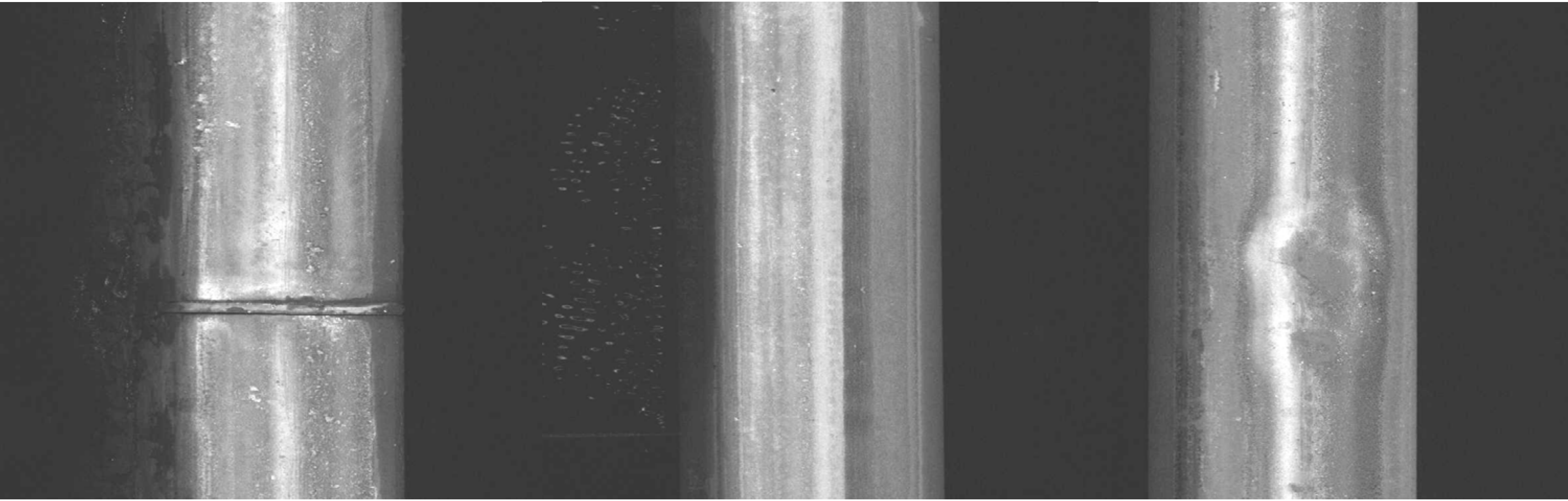
Winner 2017 and 2018

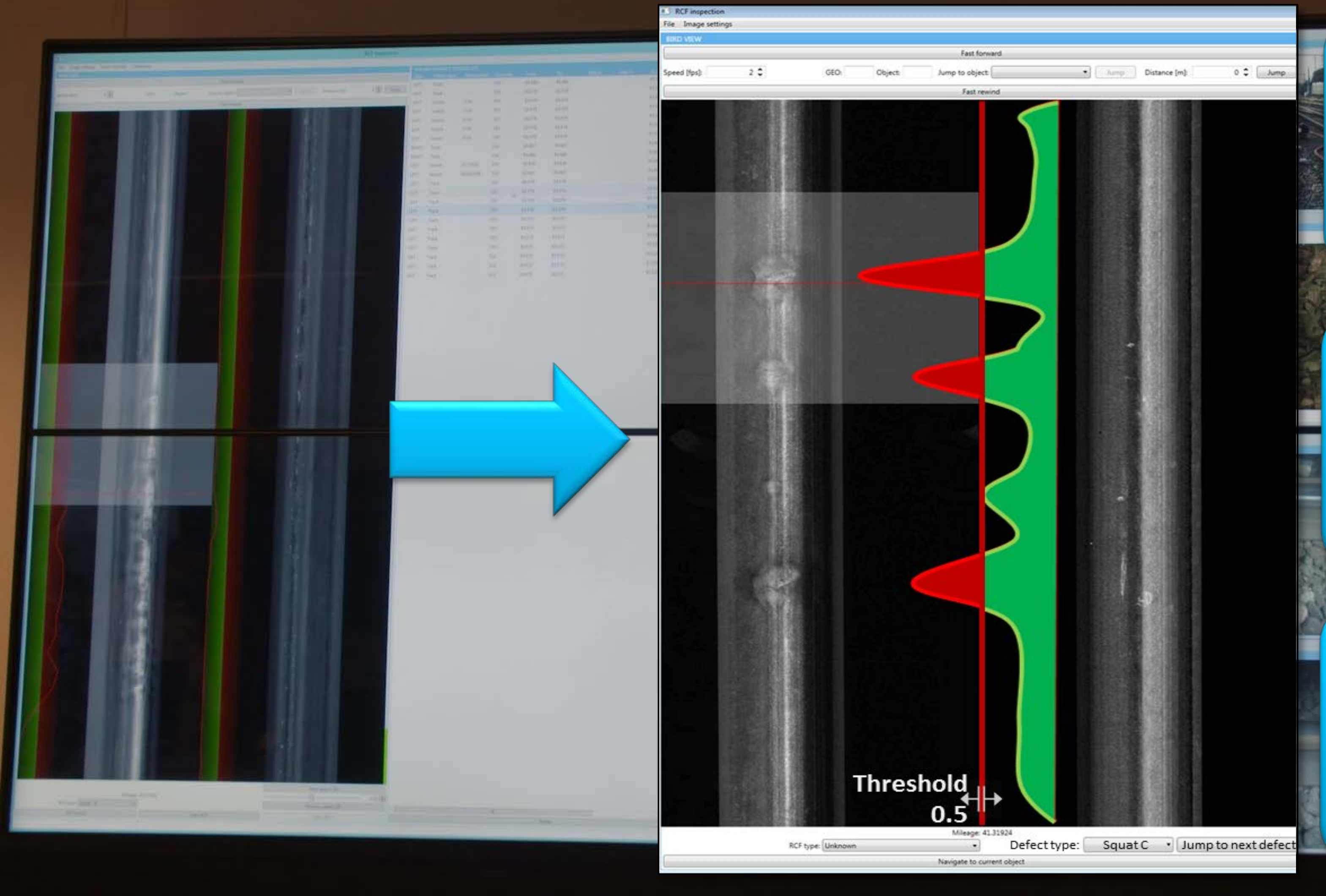
Overview of my talk

- The Project: what is it about?
- Deep Learning in a real world application.
- Transfer learning as substitute for data compression.
- Gaining trust in the model: visualising relevant pixels and regions.

Inspectation (VolkerRail) and Sherloc

Railway track inspection





500 images per second (single computer)

False negative rate < 1%

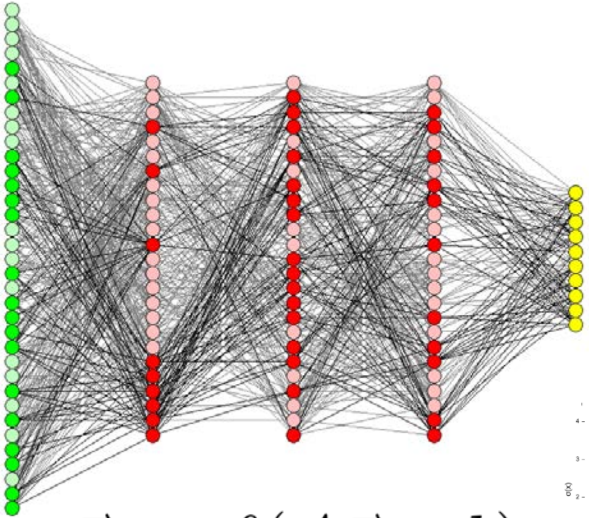
80% of work can be skipped

Overview of my talk

- ~~The Project: what is it about?~~
- Deep Learning in a real world application.
- Transfer learning as substitute for data compression.
- Gaining trust in the model: visualising relevant pixels and regions.

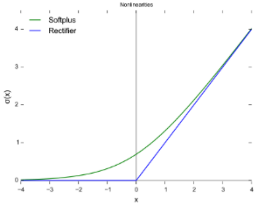
Deep Learning is Neural Networks Rebranded

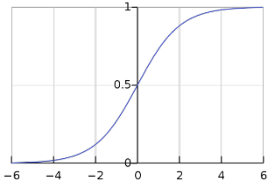
- Neural networks allow us to approximate a function assigning defect probabilities to an image.



$$\vec{x} \mapsto f(A\vec{x} + b)$$

Probability of defect: **87%**



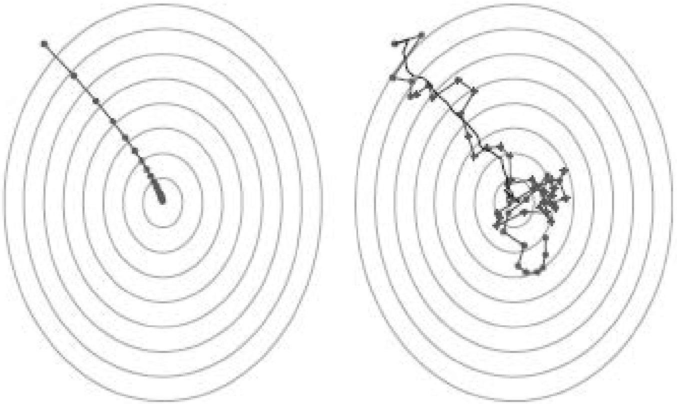


- Our neural network: 16 layers, 128 mln parameters, ReLU en Sigmoid (logistic function) activation functions.

Consultants in Quantitative Methods

Finding Good Parameter Values: Stochastic gradient descent

- Essentially Maximum Likelihood estimation with regularisation.
  - Maximum Likelihood: parameters under which the data is most likely.
  - Regularisation: techniques that improve generalisation.
- Optimisation problem!
  - Minimise the log-likelihood.
  - Algorithm: variations of stochastic gradient descent.



$$loss(\vec{\theta}) = \frac{1}{N}\sum_{i=1}^{N} y_i \log(f(\vec{x}_i|\vec{\theta})) + (1-y_i)\log(1-f(\vec{x}_i|\vec{\theta}))$$

$$reg\text{-}loss(\vec{\theta}) = loss(\vec{\theta}) + ||\vec{\theta}||_2^2$$

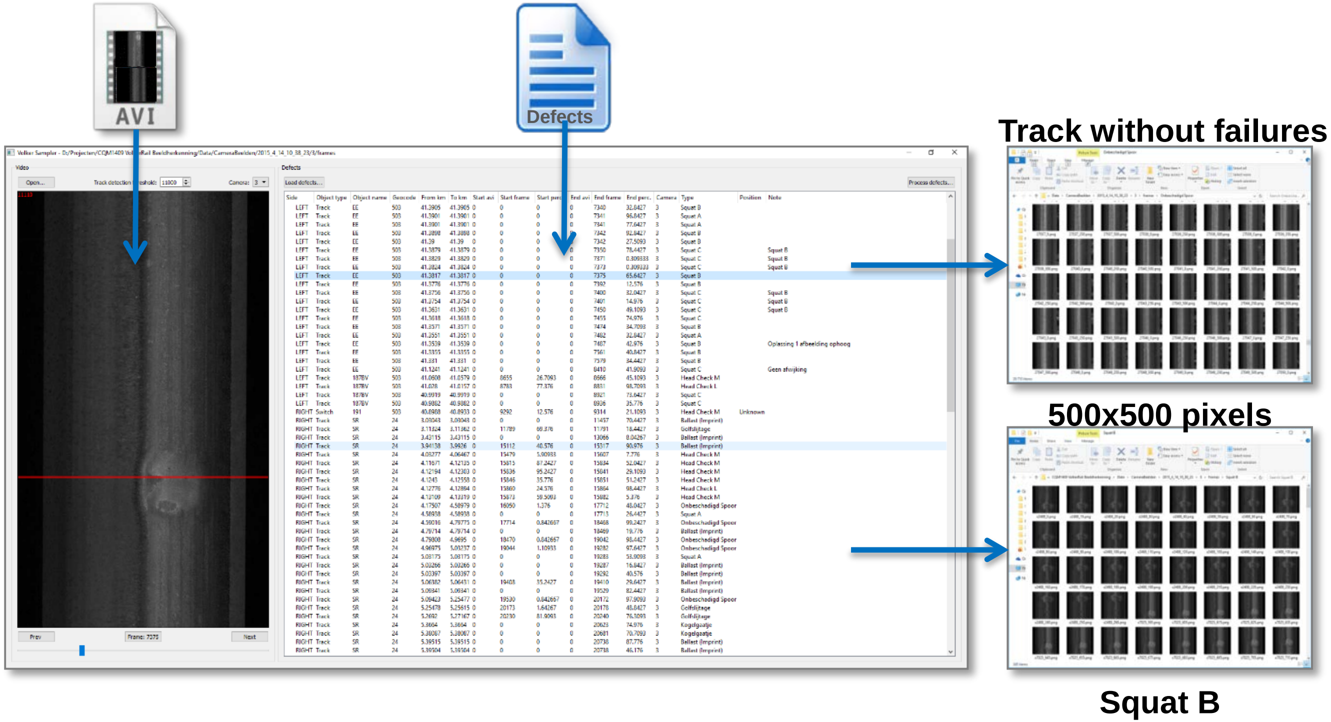
Warning: optimising the parameters of a neural network is optimising a non-convex, non-smooth, non-whatever-nice-property-you-like function.

All the Tech is Available: Thank You Gamers and Google!

TensorFlow™

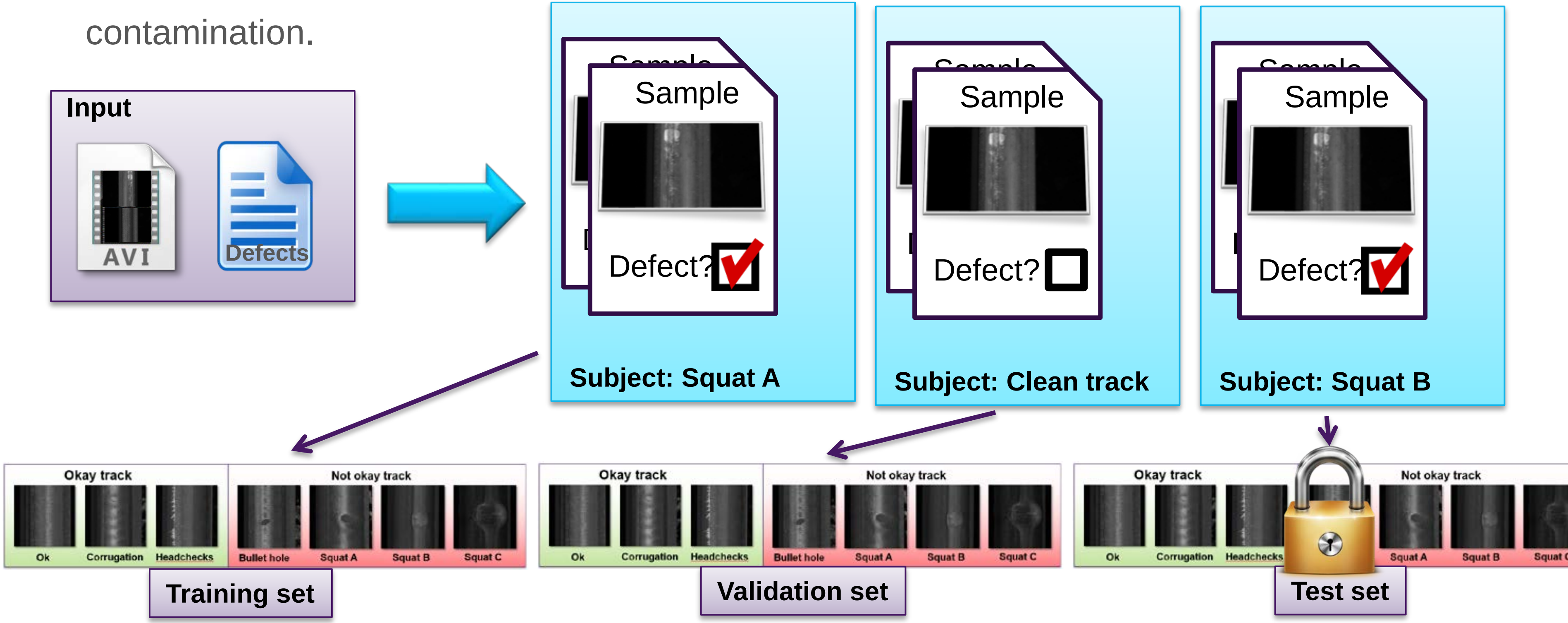
Function Approximation ≠ Real World Problem Solution

- We store a very long and narrow image as a movie file.
- Movie is 1024 pixels wide, railway track at most 400 pixels.
- Histogram based algorithm to detect the track in the image.



Function Approximation ≠ Real World Problem Solution

- Need data pipeline to get the right distribution of images, and to prevent test set contamination.



Finally: what are we optimising?

- How bad is a false-positive? (Detect a defect where there is none?)
  - And how many negative spots are there on the tracks?
- How bad is a false-negative? (Do not detect a defect while there is one?)
  - And how many positive spots are there on the tracks?

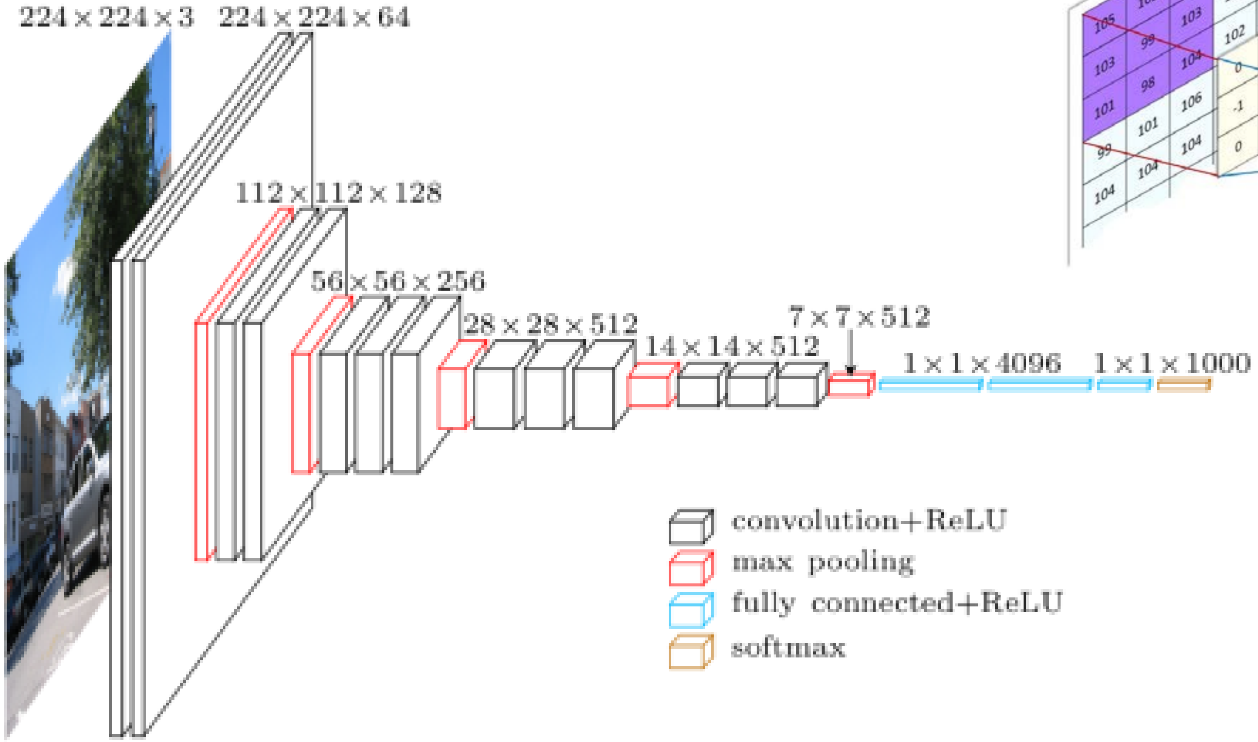
	Condition Absent	Condition Present
Negative Result	True Negative	False Negative
Positive Result	False Positive	True Positive

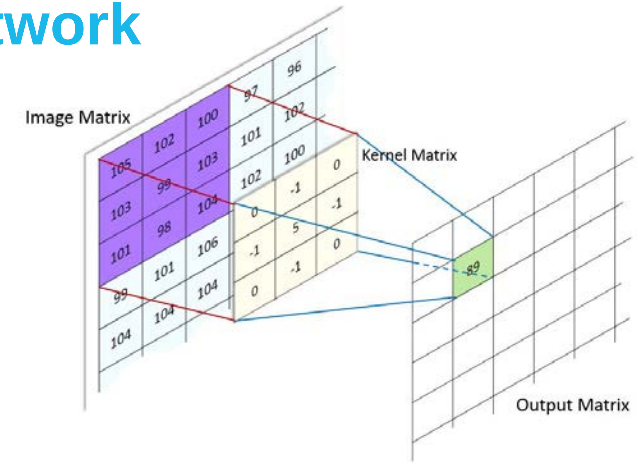
- Huge number of clear track images!
  - What if we just always classify as clean?
- Solution: weighted samples, different sliding window sampling rates, and use more data.

Overview of my talk

- ~~The Project: what is it about?~~
- ~~Deep Learning in a real world application.~~
- Transfer learning as substitute for data compression.
- Gaining trust in the model: visualising relevant pixels and regions.

Transfer learning: take someone else's network





Transfer Learning: we mostly used it for data compression.

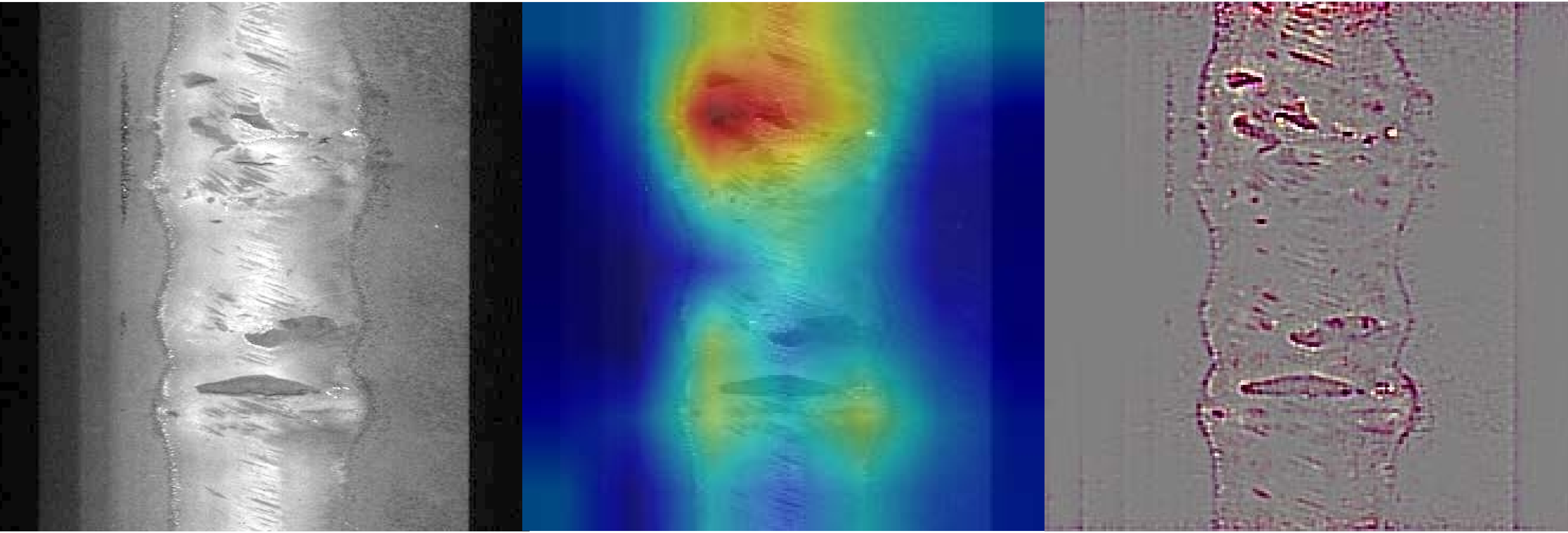
- To make our full solution work, we need a lot of data. A hell of a lot of data.
  - Why? We need to capture the all variation that can occur in the images.
  - Raindrops or Squats? Leaves on the tracks? Cold vs warm weather? Sun low at the horizon etc.
  - We are not going to manually select a representative set, and not all variation is that common.
- Many millions of 500x500 pixel images is just way too big for computer memory.
  - A lot of Disk IO and image decompressing while training.
- Our solution: take a pre-trained neural network (VGG Net)
  - Fix weights in first layers, only optimise weights in last layers.
  - 500x500 byte image (250,000 bytes) -> 1000 single precision floating point numbers (4.000 bytes).
  - Compression rate 98,4%.

Overview of my talk

- ~~The Project: what is it about?~~
- ~~Deep Learning in a real world application.~~
- ~~Transfer learning as substitute for data compression.~~
- Gaining trust in the model: visualising relevant pixels and regions.

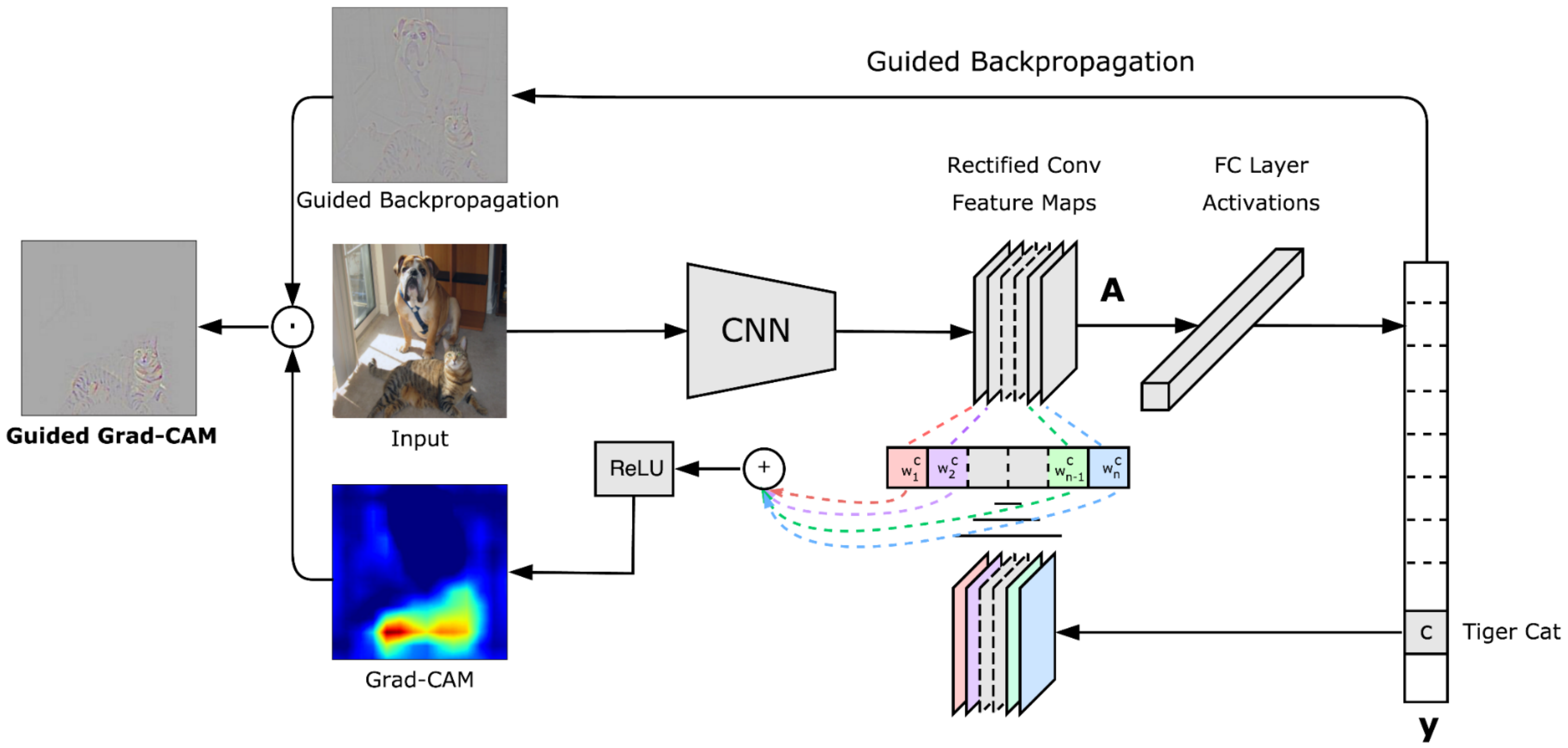
Consultants in Quantitative Methods

Demystifying the black box



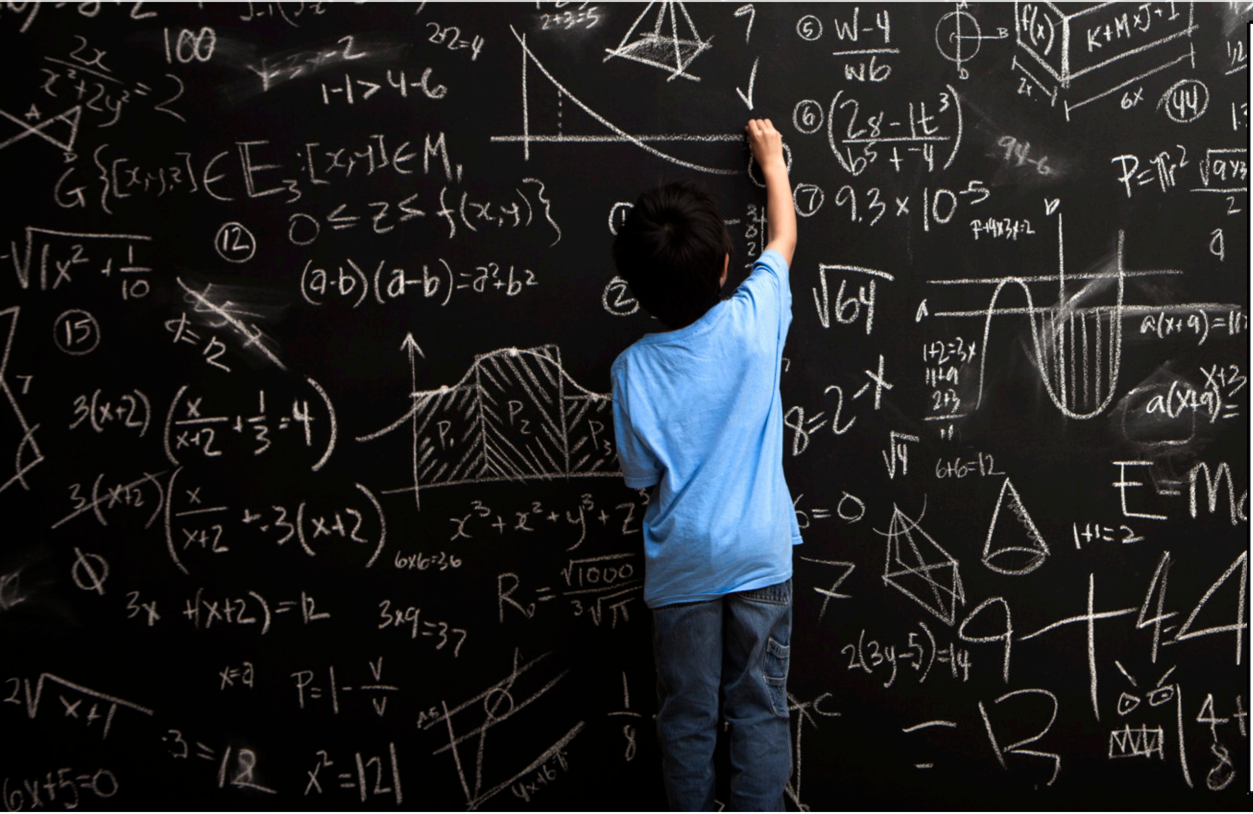
Consultants in Quantitative Methods

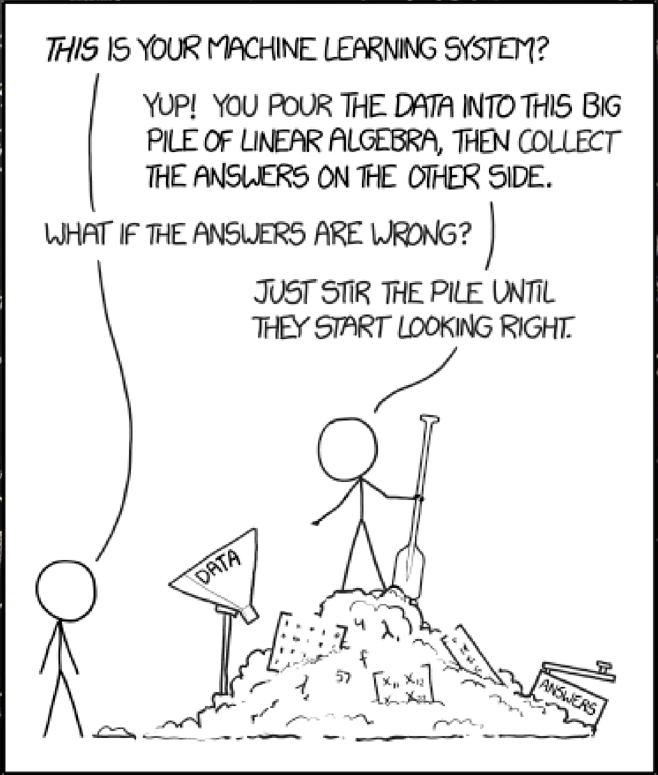
Guided Grad-CAM [Selvaraju et al.]



Consultants in Quantitative Methods

Artificial Intelligence is not magic… It's just math





To conclude… it's all about pattern recognition.

Winner 2017: Company X

Winner 2018: Company X

Winner 2019?