FFI Forsvarets forskningsinstitutt

Deep Learning on Active Sonar Data Using Bayesian Optimization for Hyperparameter Tuning

Henrik Berg
Karl Thomas Hjelmervik

ICPR 2020

Deep neural networks

- Neural networks with many layers
- Pros:
  - Powerful
  - Can discover features on their own, layer by layer
- Cons:
  - Need vast amounts of data for training
  - Prone to overfitting
  - Used to be difficult to train (and still computationally intensive)
  - **Large amount of tunable hyperparameters**

Hyperparameter Tuning

- Large number of hyperparameters
  - Number of layers, number of nodes in each layer, activation functions, dropout rates, etc
- Three different approaches:
  - Trial and Error
  - Searching the parameter space (exhaustively or randomly)
  - A more intelligent search
- Bayesian optimization:
  - Fit a function to some target function
    - (target function = ability to train a good network with given hyperparameters)
  - Initialize with random sampling
  - Iteratively improve fitting by sampling…
    - …promising areas
    - …uncertain areas

Bayesian optimization

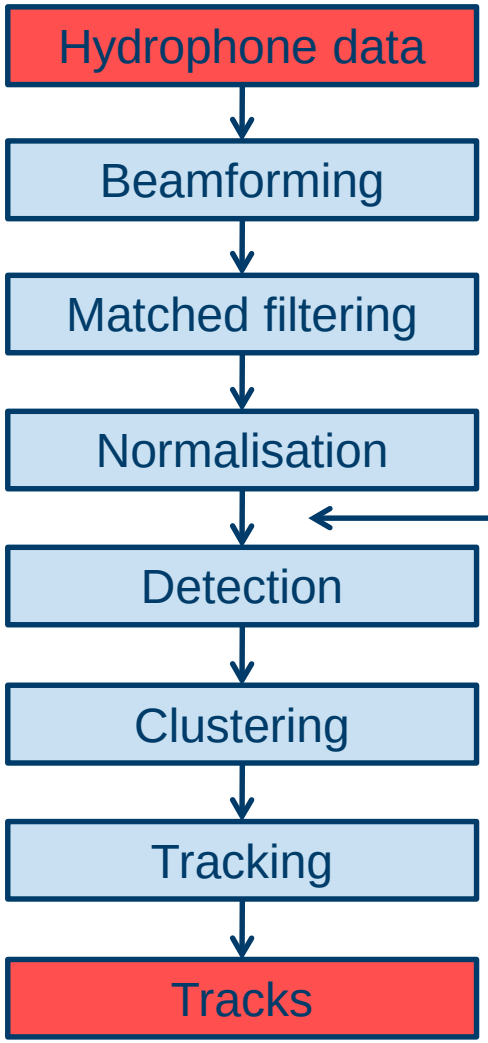
- Tested different combinations of evaluation strategies and data sets
- 36 combinations, 10 runs of Bayesian Optimization per combination
  - Total run time: approx. 4 days, 2 x 8 TITAN V GPUs
- Finally, 20 deep nets were trained for each run
- Results tested on an independent test set

Data

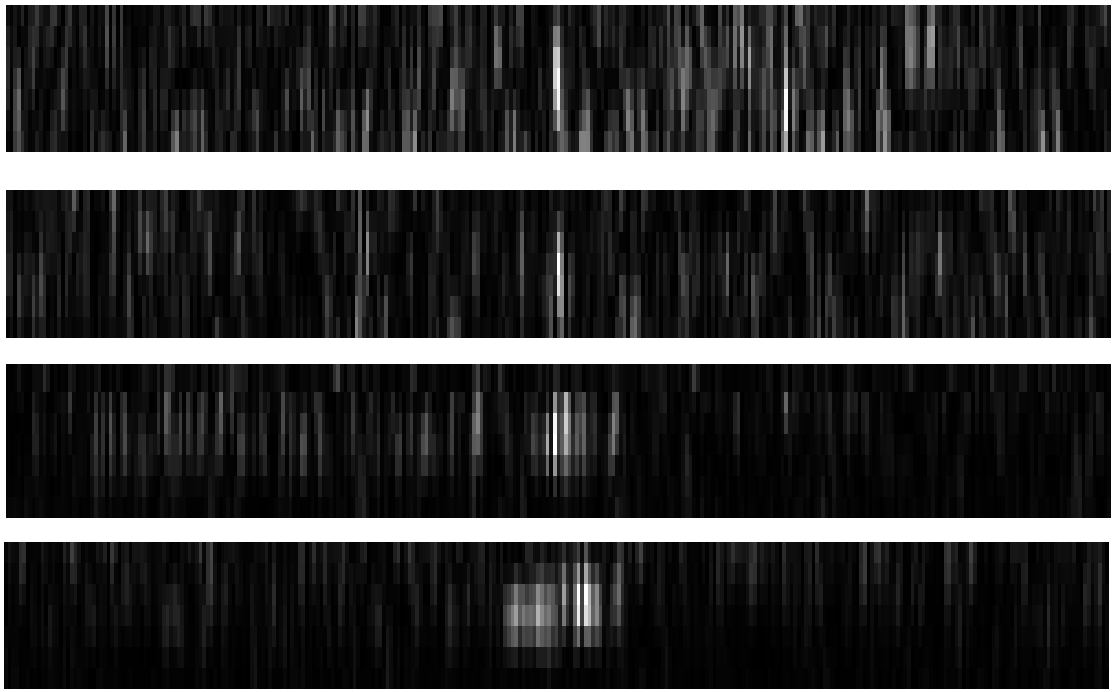
- A few thousand echoes recorded during three experiments in 2002 (PTE01, DOP03, CEX02)
- Area contained four pipelines (242 echoes)
- All echoes were semi-manually classified in advance
- Training and validation: PTE01 and CEX02
  - Three copies of each pipeline echo
  - Same number of echoes sampled from non-pipelines
  - PTE01 = 264 pipelines and 264 non-pipelines
  - CEX02 = 243 pipelines and 243 non-pipelines
- Test set: DOP03
  - Not modified in any way, and not used during training
  - 73 pipelines and 2593 non-pipelines

Processing

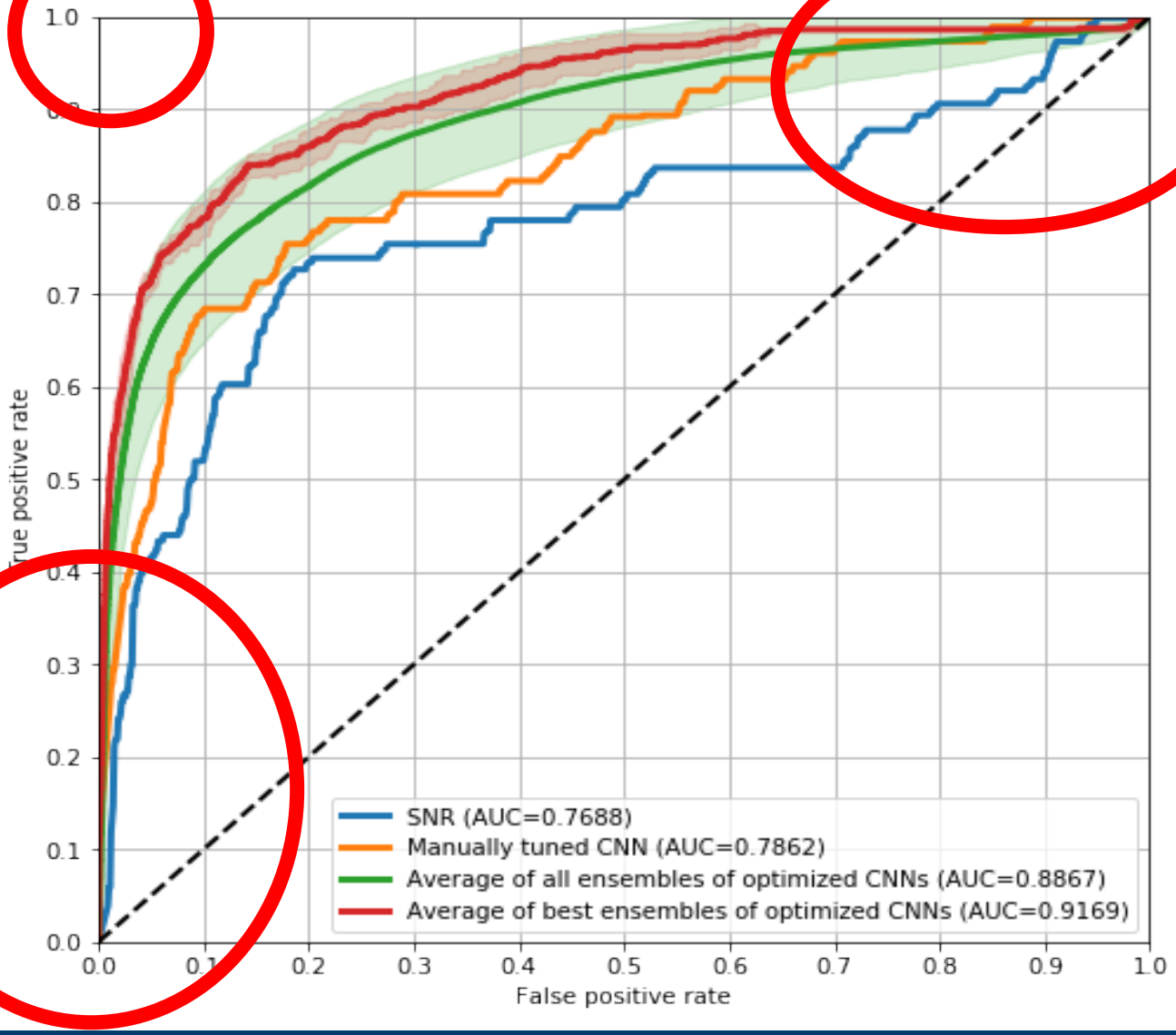
FM processing



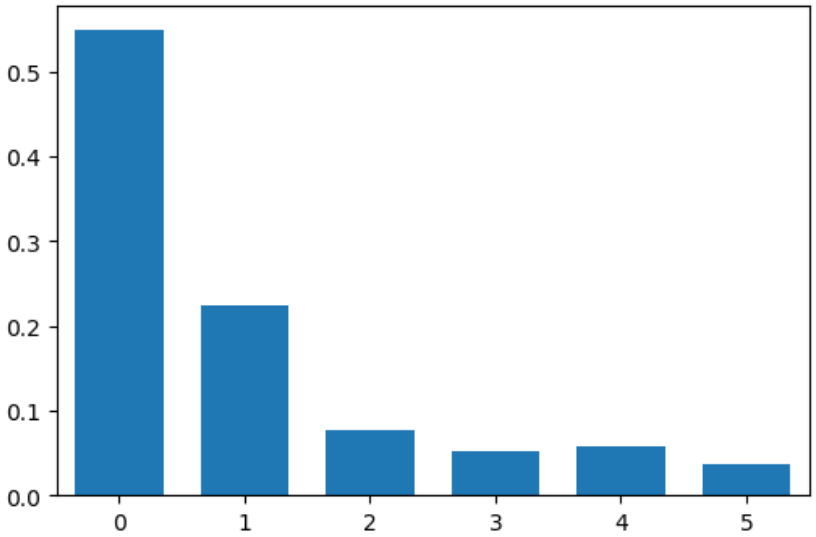
Data for training/testing



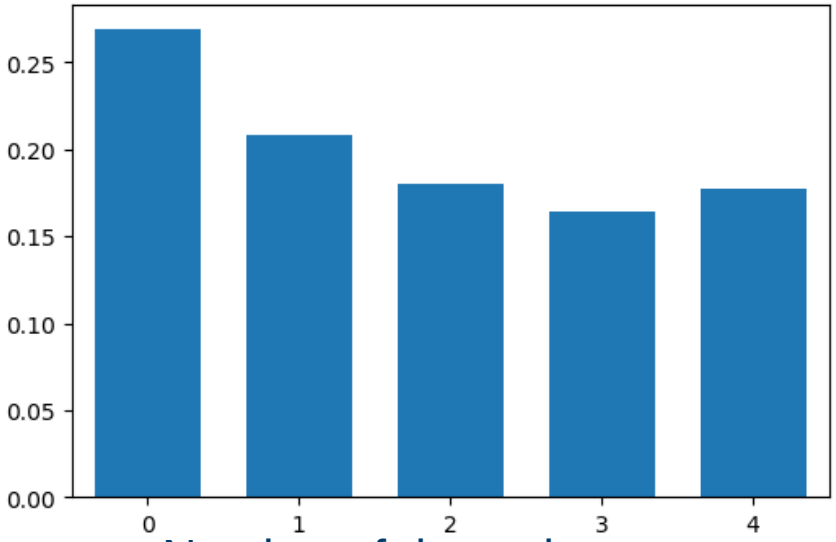
Results


1.0
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
True positive rate
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
False positive rate
SNR (AUC=0.7688)
Manually tuned CNN (AUC=0.7862)
Average of all ensembles of optimized CNNs (AUC=0.8867)
Average of best ensembles of optimized CNNs (AUC=0.9169)

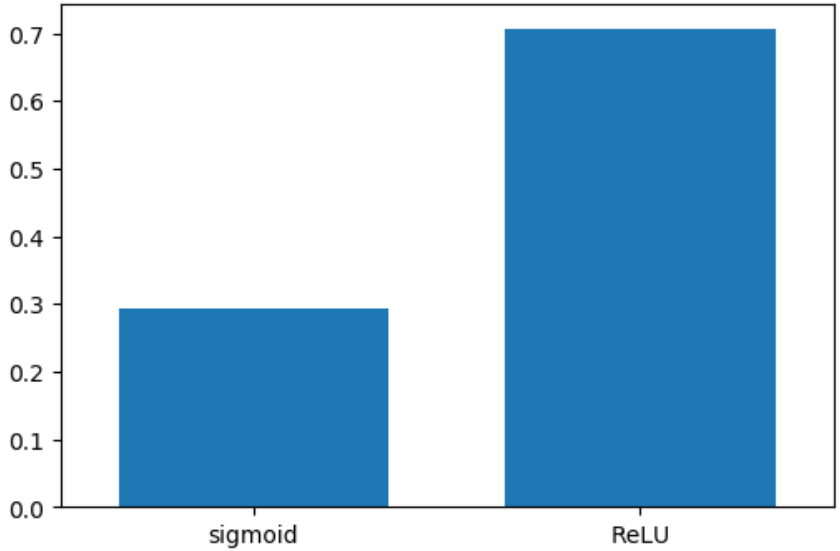
Analysing the hyperparameters



Number of convolution layers



Number of dense layers



Activation function in dense layers

Concluding remarks

- «Trial and Error» can work, but a more intelligent search can yield significantly better results
- The most sensitive parameter was choice of training set
  - PTE01 vs CEX02 – which set was most similar to the test set, DOP03?
  - Good data sets, and thorough understanding of the data, is important!