HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

Evolutionary Optimization of Neural Architectures in Remote Sensing Classification Problems

Daniel Coquelin[1], Rocco Sedona[2,3], Morris Riedel[2,3], Markus Götz[1]

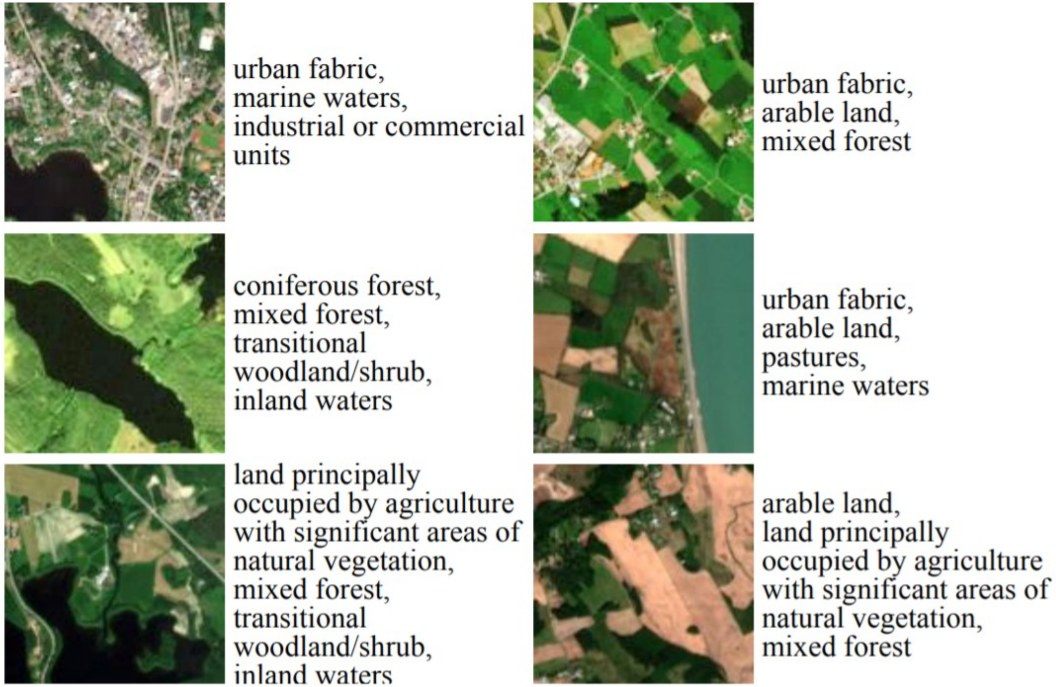
1: Steinbuch Centre for Computing (SCC), Karlsruhe Institute of Technology (KIT), Germany

2: Juelich Supercomputing Centre (JSC), Forschungszentrum Jülich (FZJ), Germany

3: School of Engineering and Natural Sciences (SENS), University of Iceland (UoI), Iceland

BigEarthNet[1]

Remote Sensing (RS) Data



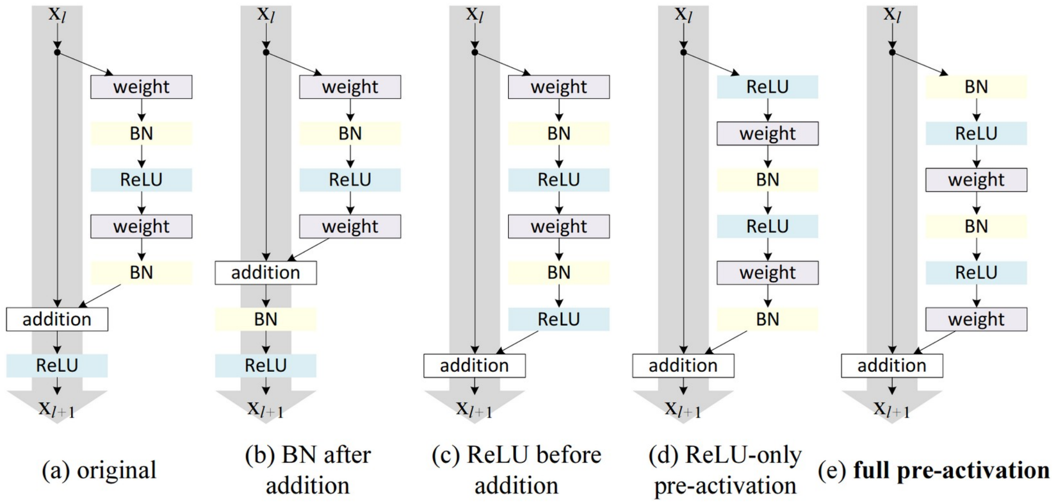
- Patches from 125 Sentinel-2 tiles (Level 2A)
- 590,326 patches, 19 labels (left)
- Each patch has 12 spectral bands
  - 4 bands at 120 x 120 px
  - 6 bands at 60 x 60 px
  - 2 bands at 20 x 20 px

1. http://bigearth.net/

Network Selection

ResNet-50

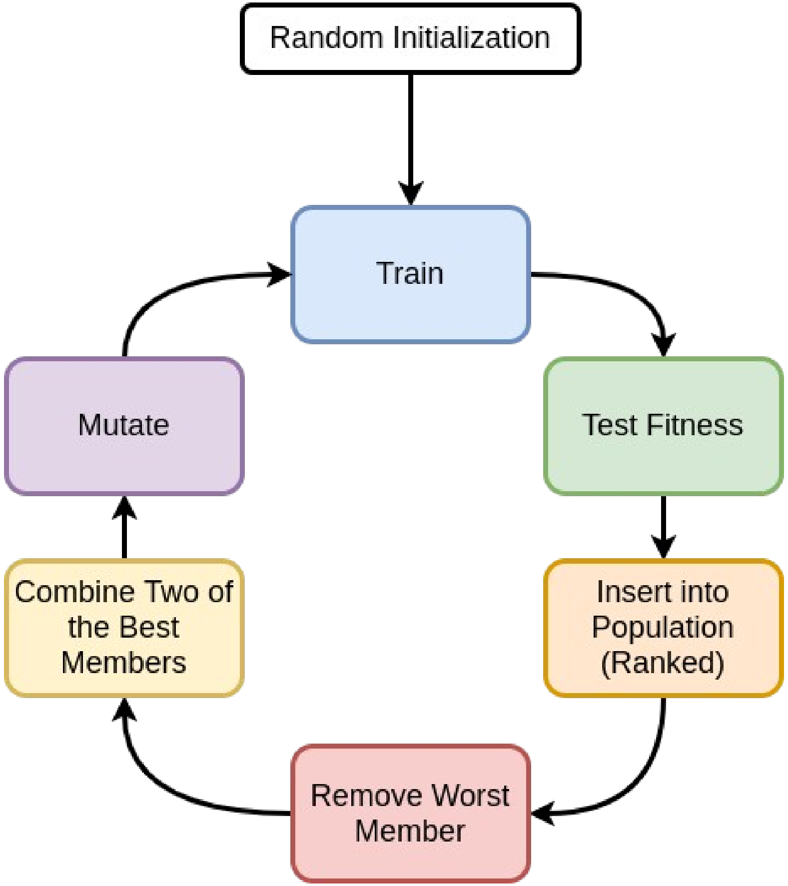
- CNN with skip connections
- “Standard” network for image tasks
- Some testing done on hyperspectral images [1]
- ResNet-50 results on BigEarthNet:
  - Micro-F_1: 77.11
  - Macro-F_1: 67.33
- What is the best configuration of the network settings (hyperparameters)?



[1] Gencer Sumbul, Jian Kang, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, et al., “BigEarthNet Dataset with A New Class-Nomenclature for Remote Sensing Image Understanding,” 2020.

Neural Architecture Search (NAS)

Evolutionary Optimization with Propulate[1]



- NAS used to find the optimal network configuration
- Evolutionary optimization:
  - Each network is a population member
  - The best networks breed to produce offspring
  - The worst networks die off
  - Over time, optimal solution is found

1. https://github.com/oskar-taubert/propulate

NAS Search Space

Network Hyperparameters

Optimizer	LR Scheduler	Activation	Activation Order	Loss
Adadelta	Exponential Decay	ELU	Original	Binary Cross Entropy
Adagrad	Inverse Time Decay	Exponential	BN after addition	Categorical Cross Entropy
Adam	Polynomial Decay	Hard sigmoid	Activation before addition	Categorical Hinge
AMSGrad		Linear	Activation-only pre-activation	Hinge
Adamax		ReLU	full pre-activation	K-L Divergence
Ftrl		SELU		Squared Hinge
Nadam		Sigmoid		
RMSprop		Softmax		
SGD		Softplus		
		Softsign		
		Swish		
		tanh		

NAS Experimental Setup

- Network (ResNet-50) implemented in TensorFlow
- No image augmentations
- Trained on multiple A100 GPUs at Horeka at KIT
- Early stopping when validation scores are not improving for 10 epochs
- Individual NASs run for Adam, Adamax, Nadam, and RMSprop.
  - NaN values appear if hyperparameters are unfavorable
  - If NaNs appear, the optimizers will be removed from the search space quickly
    - This is unfair as they are removed for stability and not accuracy

Experimental Results

Most Accurate Tested Network

- Most accurate network:
  micro-F1: 77.25
  macro-F1: 69.57
  - Adadelta, polynomial decay LR, softmax activation, binary cross entropy loss, full pre-activation, 128 filters
- **Epochs trained: 10** (previous was 100)
- Most classes show better performance

Class	Found	Original
Urban fabric	**75.86**	74.84
Industrial or...	45.79	**48.55**
Arable land	**84.82**	83.85
Permanent crops	**59.63**	51.91
Pastures	**74.54**	72.38
Complex cultivation...	**69.00**	66.03
Land principally...	**65.37**	60.94
Agro-forestry areas	**77.35**	70.49
Broad-leaved forest	**77.27**	74.05
Coniferous forest	**86.25**	85.41
Mixed forest	**82.15**	79.44
Natural grassland...	**47.77**	47.55
Moors, heathland...	**64.46**	59.41
Transitional woodland...	**64.72**	53.47
Beaches, dunes, sands	44.09	**61.46**
Inland wetlands	**61.99**	60.64
Coastal wetlands	**57.31**	47.71
Inland waters	**85.53**	83.69
Marine waters	**97.93**	97.53

Conclusion

Further Research

- Evolutionary NAS found a more accurate network using better hyperparameters
- Network results are dependent on the search space
- Higher accuracy may require specialized network architectures for hyperspectral data

Shameless Plug

Check out my other remote sensing adjacent work!

- Large scale, distributed data analysis:
  - https://github.com/helmholtz-analytics/heat/
- Hyperspectral denoising in Python on GPUs:
  - https://github.com/Helmholtz-AI-Energy/HyDe