Evolution of Convolutional Neural Networks

Alexander Polok

Brno University of Technology, Faculty of Information Technology
Božetěchova 1/2. 612 66 Brno - Královo Pole

xpolok03@fit.vutbr.cz

BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

May 8, 2022

No prior knowlege, let evolution handle it!

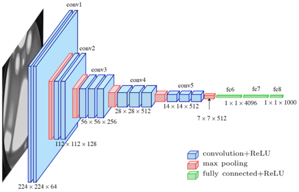


Figure: VGG16 architecture (1)

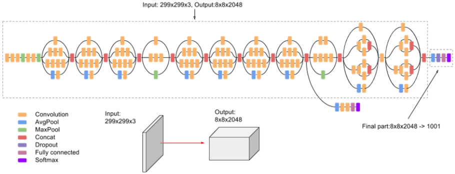


Figure: Inception v3 architecture (2)

MNIST vs. FashionMNIST

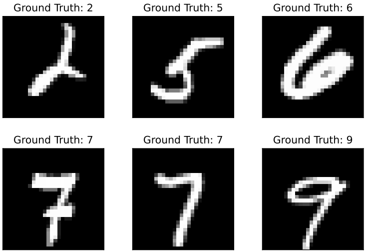


Figure: MNIST dataset
SOTA accuracy $\sim$ 99.9%,
CNN with 2 layers $\sim$ 98%

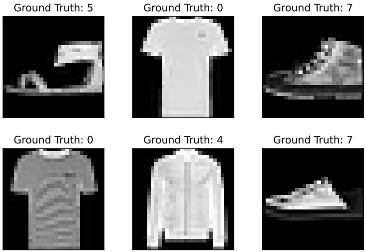


Figure: Fashion MNIST dataset
SOTA accuracy $\sim$ 96%,
CNN with 2 layers $\sim$ 83%

Hyperparameters

- Number of epochs[1]
- Kernel size
- Number of output channels
- Genotype size
- Evolutionary alghorithm and it's parameters
- Set of primitive operations
- Architecture type

[1] Adam optimizer (lr=0.001, betas=(0.9, 0.999))

- None
- Identity
- Conv2d 1×1, BatchNorm2d, ReLU
- Conv2d 1×1, BatchNorm2d, ELU
- Conv2d $C \times C$, BatchNorm2d, ReLU
- Conv2d $C \times C$, BatchNorm2d, ELU
- Dropout2d
- MaxPool2d $C \times C$
- AvgPool2d $C \times C$

All primitives are of stride one and the convolved feature maps are padded to preserve their spatial resolution.

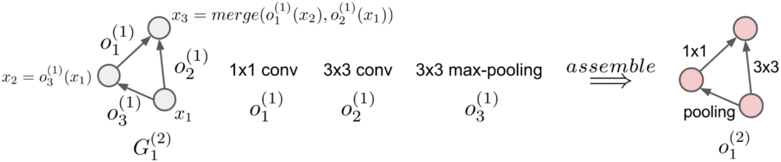


Figure: Flat representation (level 2 motif) - NN is constructed from set of primitive operations (level 1 motifs) (3).

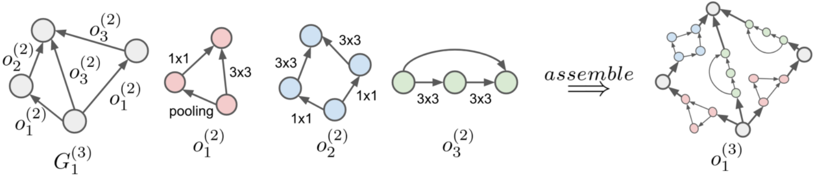


Figure: Hierarchical representation - NN is construted recursively from lower level motifs (3).

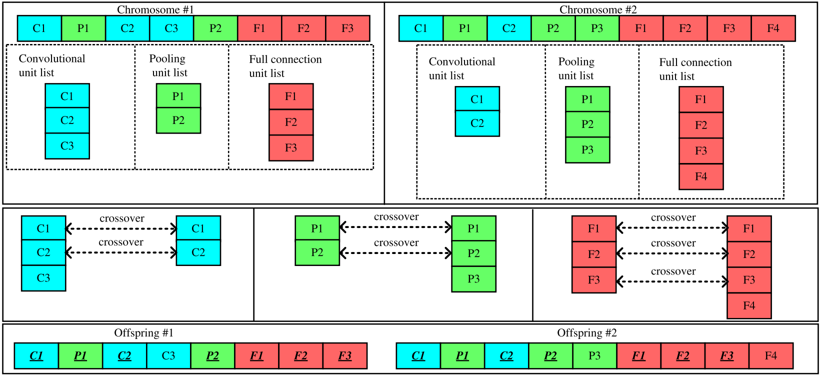


Figure: Variable lenght representation (4)

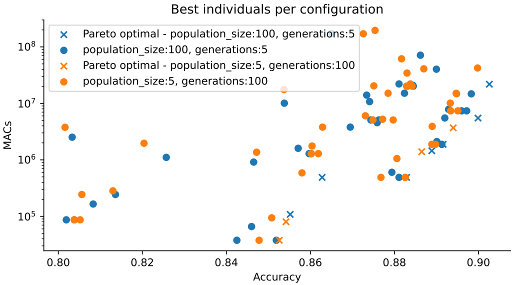
- Flat - 96 experiments
  - population size, number of generations - $100 \times 5, 5 \times 100$[2]
  - kernel size - 3, 5
  - channels - 2, 8, 16, 32
  - number of nodes - 2, 4, 8
  - init mutations - 0, 1000
- Hierarchical - 40 experiments 100×5 (pop,gen)
  - channels - 2, 8, 16, 32
  - kernel size - 3, 5
  - number of nodes - '2,2,2;2', '2,2,2;2,2,2;2', '3,3,3;3,3,3;3', '4,4,4;4', '3,3,3;6'
- Variable length - 40 experiments, 50×10 (pop,gen)[3]
  - channels - 2, 8, 16, 32
  - kernel size - 3, 5
  - number of nodes - '2,1', '3,2', '5,2', '7,2', '5,3'

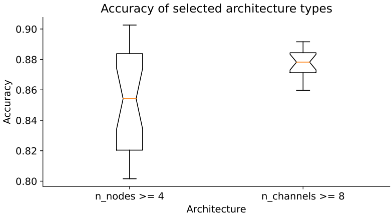
Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.

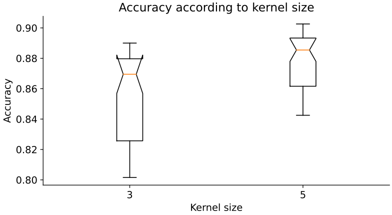
[2] 7 000 individuals were evaluated in original paper (3).

[3] 10 000 individuals were evaluated in original paper (4).

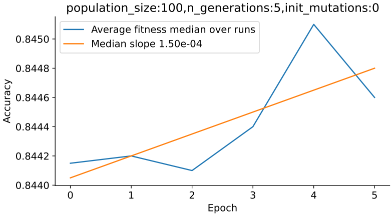
Flat - population size versus number of generations, kernel size, architecture

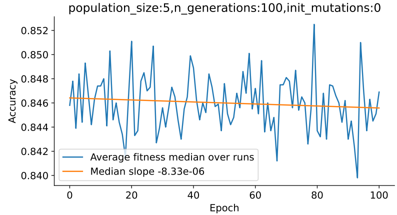


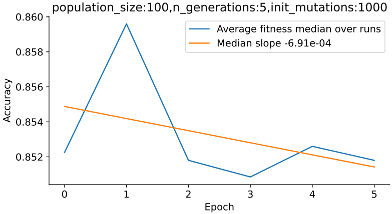


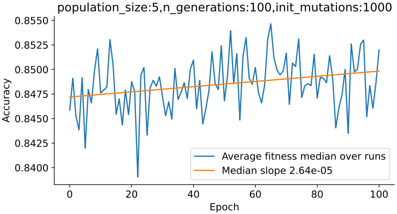


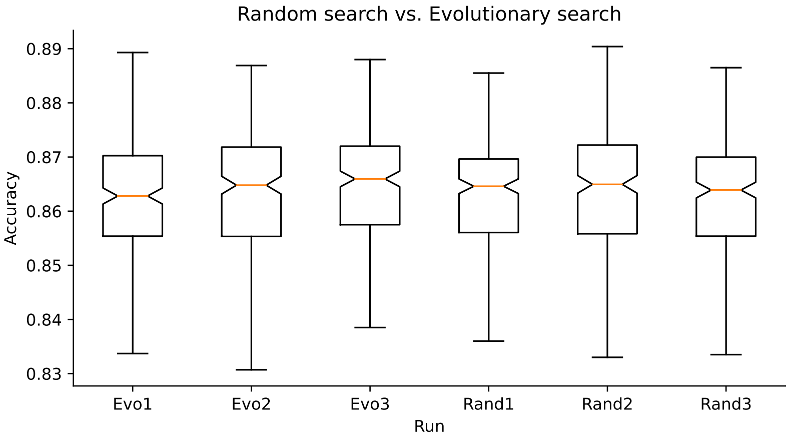
Flat - median fitness through epochs



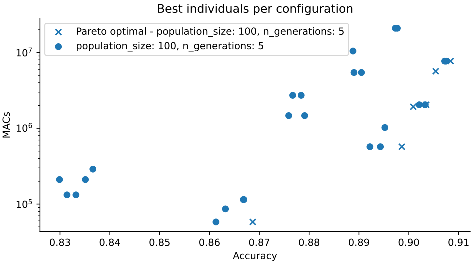


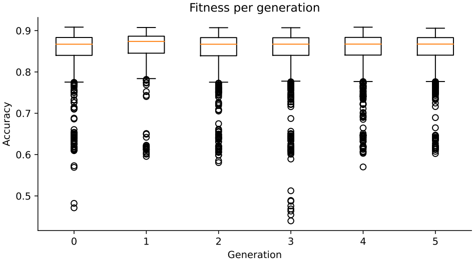




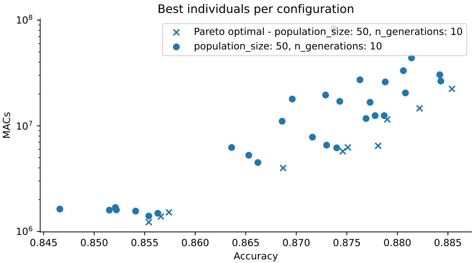

Random search vs. Evolutionary search
Accuracy
0.89
0.88
0.87
0.86
0.85
0.84
0.83
Evo1
Evo2
Evo3
Rand1
Rand2
Rand3
Run

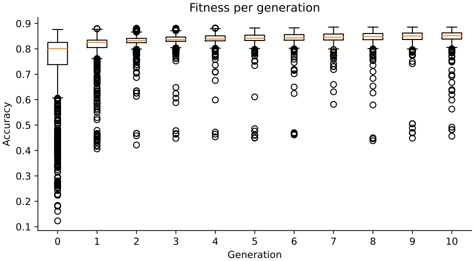
Hierarchical representation

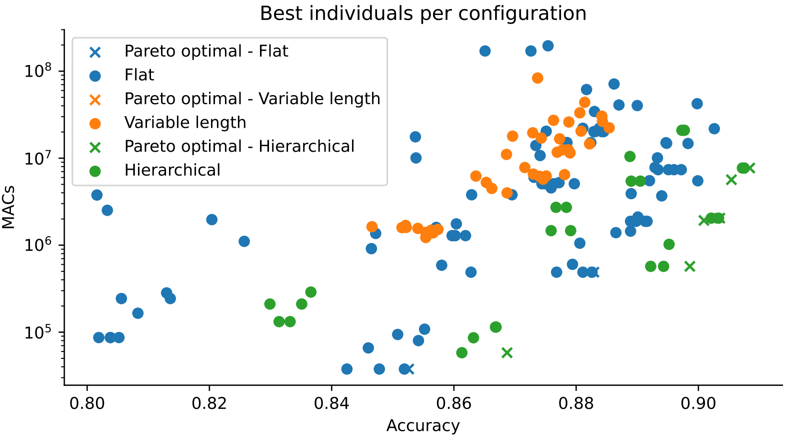




Variable length representation






Best individuals per configuration
Pareto optimal - Flat
Flat
Pareto optimal - Variable length
Variable length
Pareto optimal - Hierarchical
Hierarchical
MACs
10^8
10^7
10^6
10^5
0.80
0.82
0.84
0.86
0.88
0.90
Accuracy

- Max Ferguson, Ronay ak, Yung-Tsun Lee, and Kincho Law.
  Automatic localization of casting defects with convolutional neural networks.
  pages 1726–1735, 12 2017.

- Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna.
  Rethinking the inception architecture for computer vision.
  CoRR, abs/1512.00567, 2015.

- Hanxiao Liu, Karen Simonyan, Oriol Vinyals, Chrisantha Fernando, and Koray Kavukcuoglu.
  Hierarchical representations for efficient architecture search.
  CoRR, abs/1711.00436, 2017.

References II

Yanan Sun, Bing Xue, Mengjie Zhang, and Gary G. Yen.
Evolving deep convolutional neural networks for image classification.
IEEE Transactions on Evolutionary Computation, 24(2):394–407, 2020.