Multi-objective Bayesian Optimization for Neural Architecture Search

Petra Vidnerová, Jan Kalina

Institute of Computer Science
The Czech Academy of Sciences

ICAISC 2022

Motivation

Neural Architecture Search (NAS)

- hot research topic nowadays
- need for automatic design of neural architectures

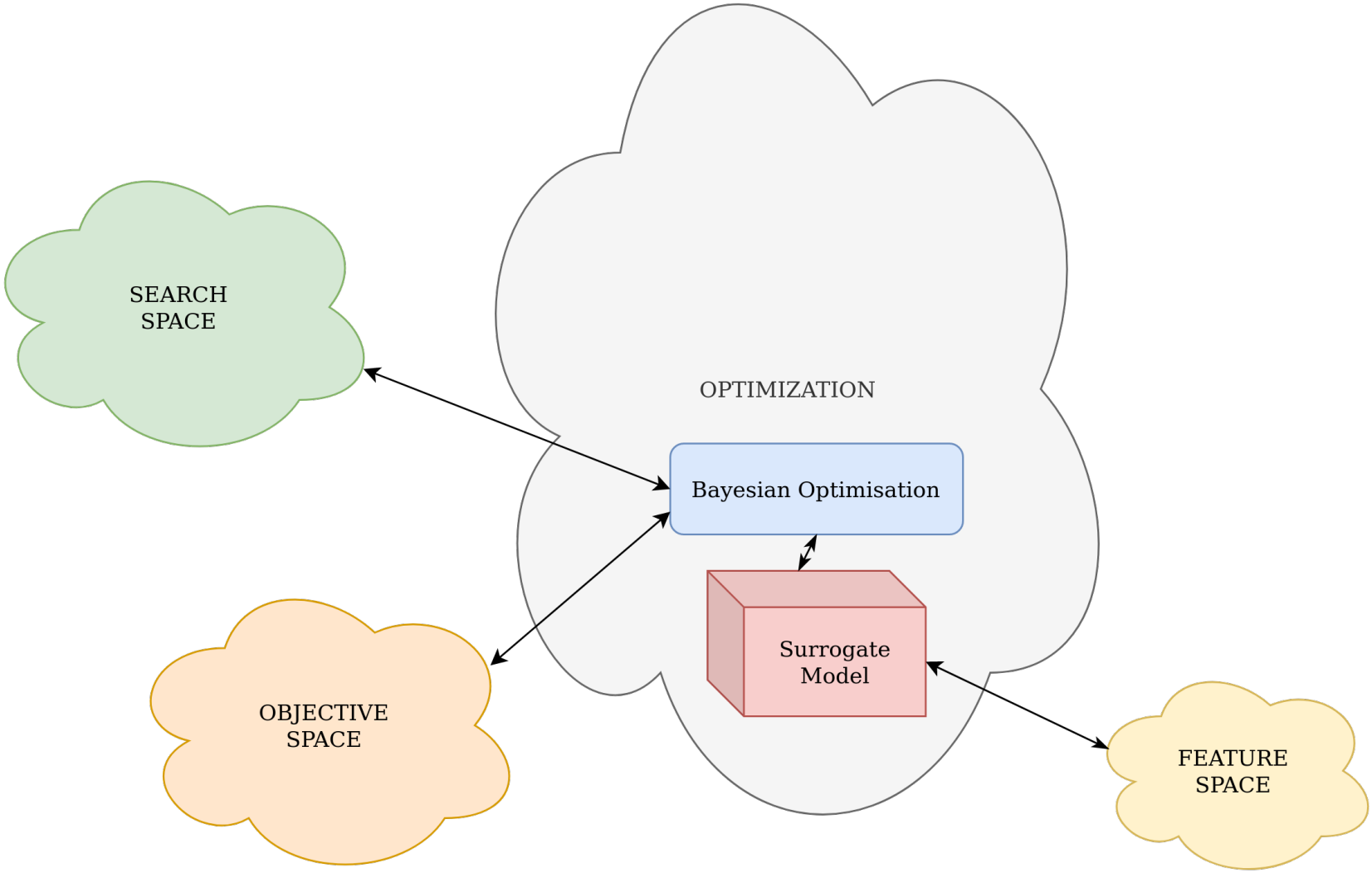
Bayesian Optimisation

- suitable for problems with computationally expensive objective functions

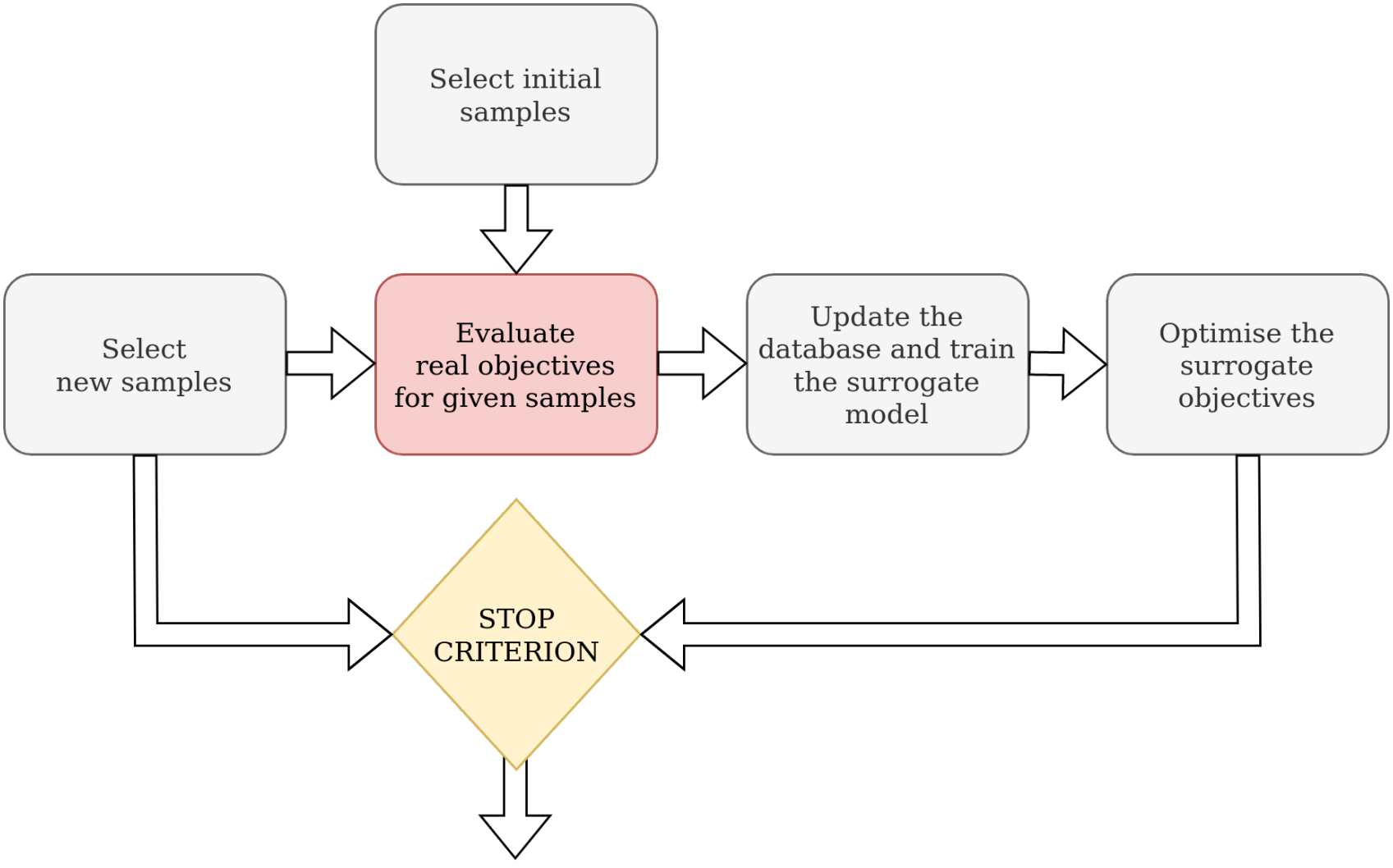
Multi-objective Optimisation

- optimising not only network performance, but also size, energy consumption, etc.

Multi-objective Bayesian Optimization for NAS

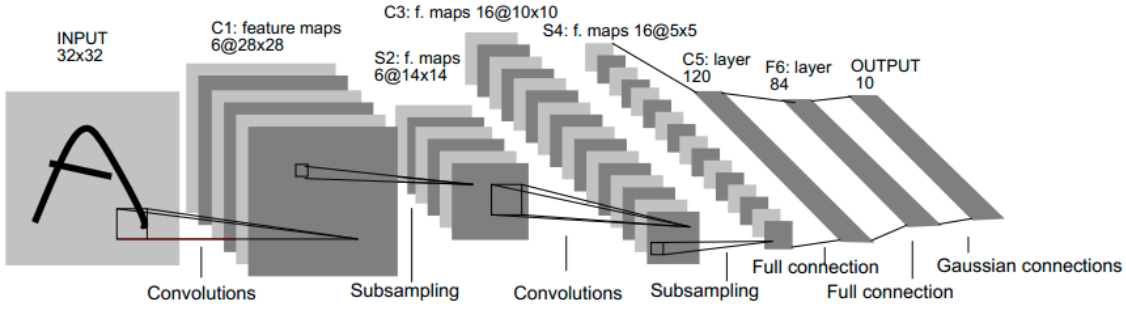


Bayesian Optimization



Search Space

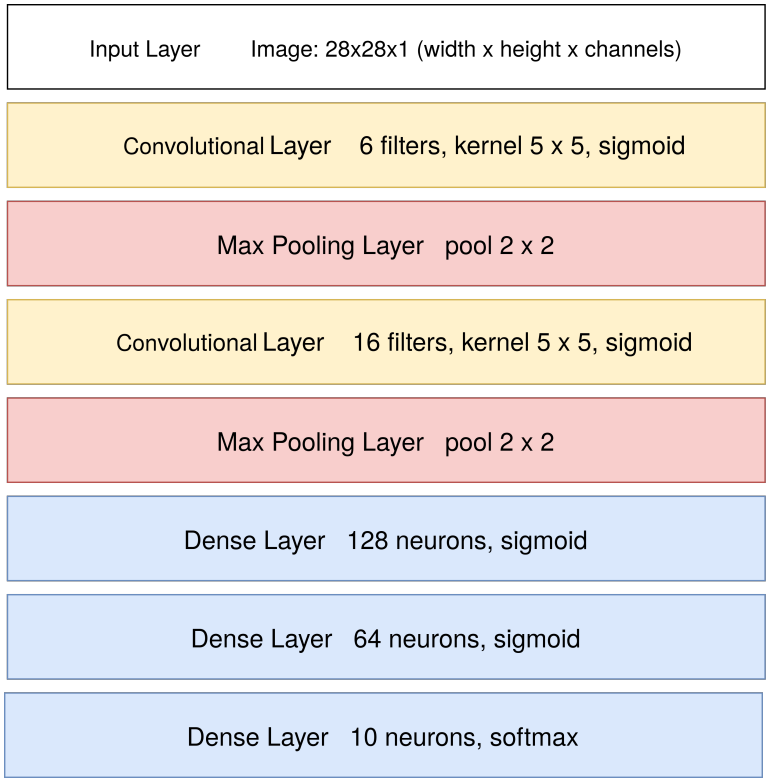
Feed-forward dense and convolutional networks



Individual encoding

- individuals consists of two parts: convolutional and dense
- each part consists of blocks corresponding to layers
- block is a tuple coding parameters of a layer

Search Space - Individual Encoding



Code:

```
[[("Conv", 6, 5, "sigmoid"),
  ("Pool", 2),
  ("Conv", 16, 5, "sigmoid"
  ("Pool", 2)],
 [(128, "sigmoid",0.0),
  (64, "sigmoid", 0.0)]]
```

Building Blocks

Optimising the surrogate function

- NSGAII algorithm
- using crossover on whole blocks, mutations modify parameters of blocks

Surrogate model

- Gaussian processes with Matern kernel
- works on feature space

Feature space

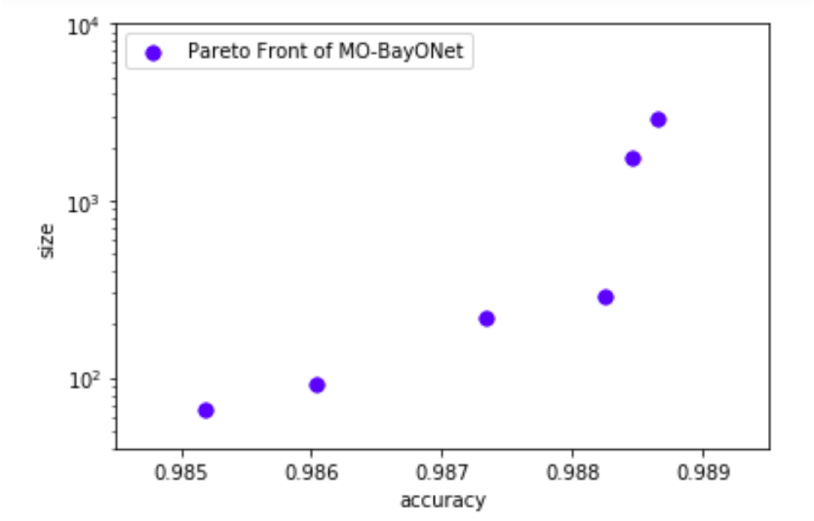
- inputs for surrogate model, fixed size vector
- features describing the network

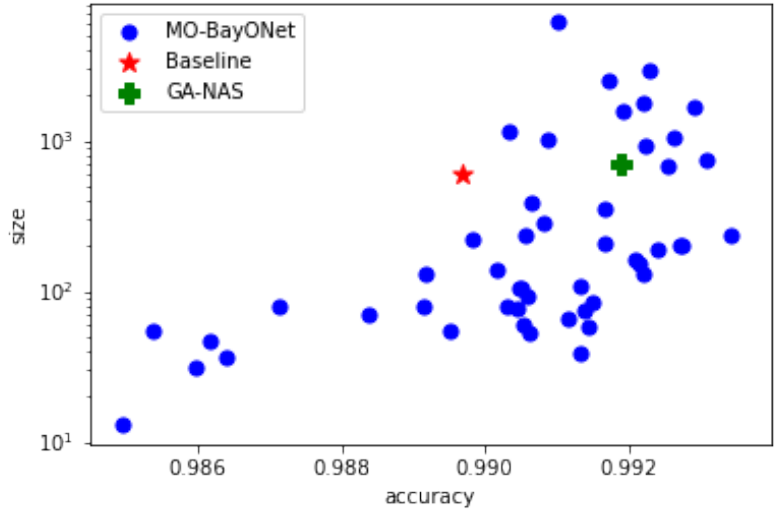
N_A	The number of network parameters
N	The number of convolutional layers
P	The number of pooling layers
$\bar{k}$	The mean size of the convolutional filter
a_{at}^{C}	Relative numbers of individual activations in convolutional part
M	The number of dense layers
a_{at}^{D}	Relative numbers of individual activations in dense part
d	The minimal, maximal, and mean dropout values

Experiments

- **Datasets:** MNIST, fashion-MNIST, CIFAR10
- **Objectives:** cross-validation accuracy, network size (number of parameters)
- **Iterations:** 100
- **Final metric:** accuracy on test set (after 20 epochs)
- **Compared to:** small baseline model, solution found by GA

Experimental Results





Thank you! Questions?