STOR566: Introduction to Deep Learning

Lecture 20: Neural Architecture Search

Yao Li
UNC Chapel Hill

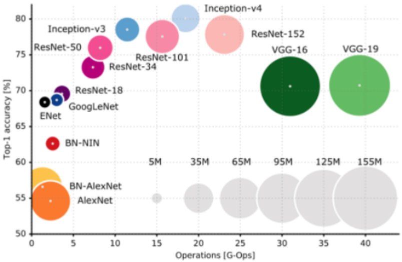
Nov 8, 2022

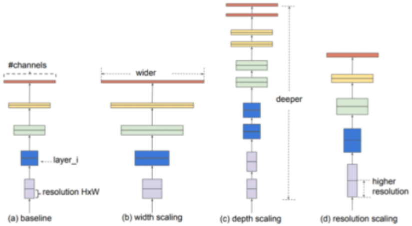
Materials are from *Deep Learning (UCLA)*

Neural Architecture Search (NAS)

NAS: Background

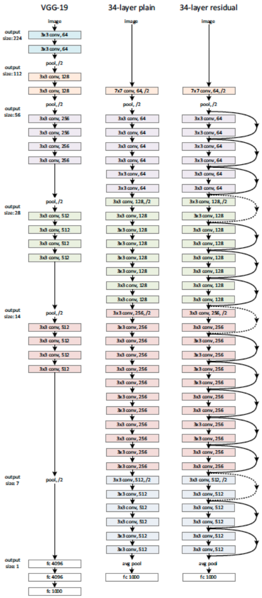
- Neural network architecture is important for both accuracy and efficiency



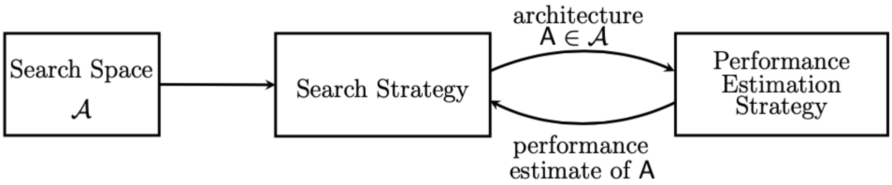


NAS: Why this Architecture?

- Architecture of VGG19 and ResNet34 on ImageNet
- How does people come up with this final architecture?
- Can we automatically design an architecture?



NAS: Overview

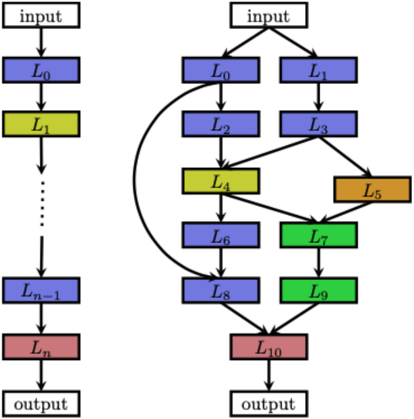


Picture from Elsken et al., Neural Architecture Search: A Survey, JMLR, 2019

- Abstract illustration of Neural Architecture Search methods.

Search Space

- An illustration of different architecture spaces.
- Left: Chain-structured neural networks
- Right: A more complex search space with additional layer types



Picture from Elsken et al., Neural Architecture Search: A Survey, JMLR, 2019

Search Space: Parameters

For Chain-structured neural networks:

- n: the (maximum) number of layers
- the type of operation in a layer, e.g., pooling, convolution, full-connected, etc.
- hyper-parameters associated with the operations

For more complex search space:

- Skip connection
- Dense net connection
- ...

Search Space: Parameters

For Chain-structured neural networks:

- n: the (maximum) number of layers
- the type of operation in a layer, e.g., pooling, convolution, full-connected, etc.
- hyper-parameters associated with the operations

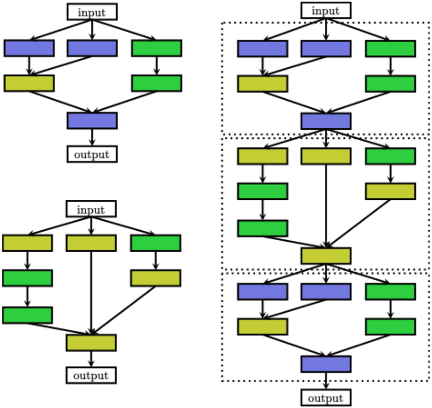
For more complex search space:

- Skip connection
- Dense net connection
- ...

The optimization problem can be very hard!

Search Space: Cell-based

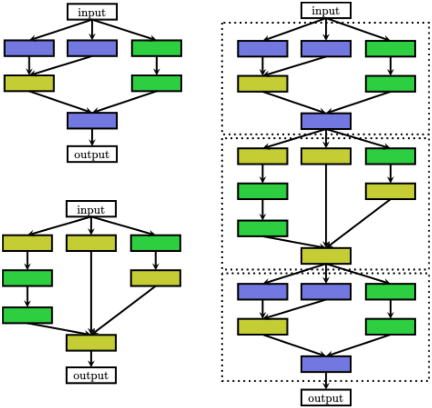
- Search the architecture for a cell (block)
- Build final architecture by stacking cells.
- Zoph et al., Neural architecture search with reinforcement learning. ICLR, 2017.: normal cell, reduction cell



Picture from Elsken et al., Neural Architecture Search: A Survey, JMLR, 2019

Search Space: Cell-based

- Search the architecture for a cell (block)
- Build final architecture by stacking cells.
- Zoph et al., Neural architecture search with reinforcement learning. ICLR, 2017.: normal cell, reduction cell



Picture from Elsken et al., Neural Architecture Search: A Survey, JMLR, 2019

How to determine the macro-architecture?

Search Strategy

Overview of Search Strategies:

- Bayesian optimization
- Evolutionary methods
- Reinforcementt learning (RL)
- Gradient-based methods
- ...

Search Strategy: RL

In 2016, Reinforcement learning (RL) is proposed for NAS

- A better (structured) representation of search space
- Learning a controller to generate architectures

Search Strategy: RL

In 2016, Reinforcement learning (RL) is proposed for NAS

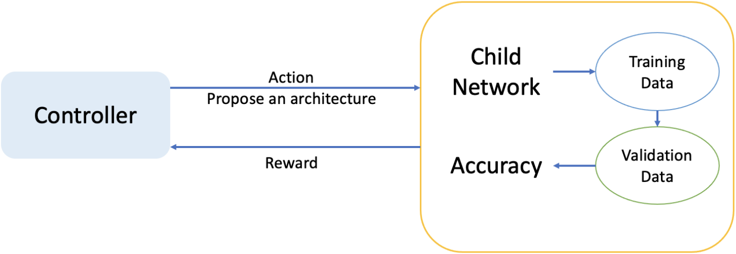
- A better (structured) representation of search space
- Learning a controller to generate architectures

Papers:

- Zoph et al., Neural Architecture Search with Reinforcement Learning. ICLR, 2017.
- Baker et al., Designing Neural Network Architectures using Reinforcement Learning. ICLR, 2017.

Architecture	Test Error (%)	Search Cost (GPU days)	Search Method
ResNet (He et al., 2016)	4.62	-	manual
DenseNet-BC (Huang et al., 2017)	3.46	-	manual
NAS-RL (Zoph & Le, 2017)	3.65	22,400	RL

NAS with RL



- Controller: a RNN to propose an architecture
- Train and evaluate proposed architecture
- Update the controller with the reward

Controller

- Controller to generate hyper-parameters of neural networks
- RNN as backbone
- Simple example: generate hyper-parameters of a chain-structured CNN

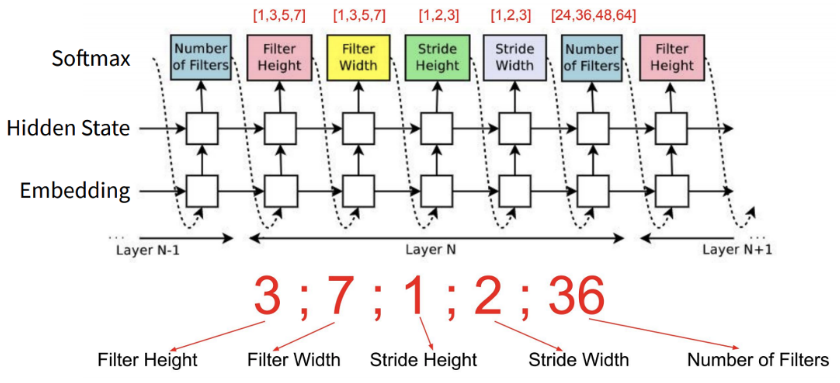
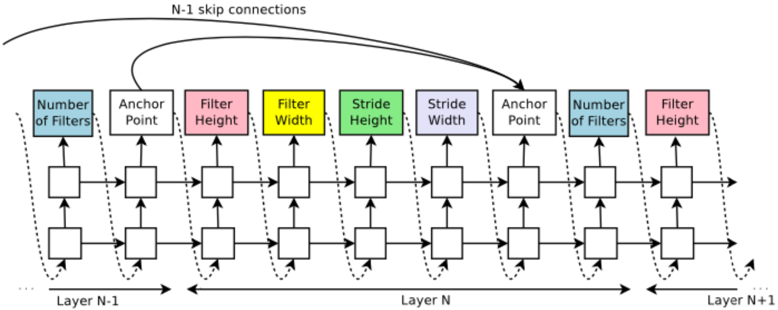


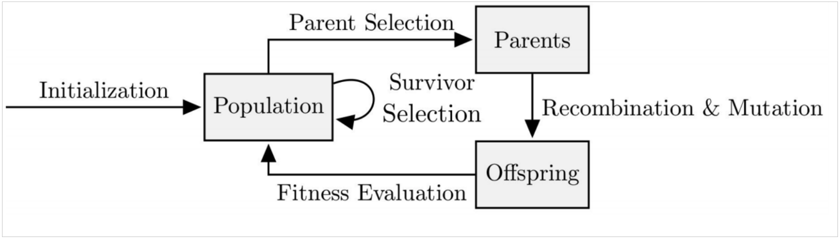
Figure from Zoph et al., Neural Architecture Search with Reinforcement Learning. ICLR, 2017.

Controller: Anchor Point

- Anchor point to form skip connection
- At layer N, $N-1$ sigmoids to indicate the previous connections



Evolutionary Algorithm

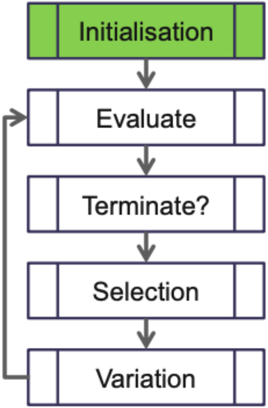


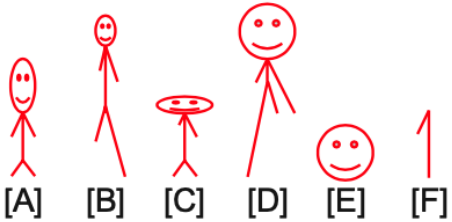
- A nature inspired approach to optimization
- Process of getting the most out of something
- Inspired by the notion of survival of the fittest from Darwinian Evolution and modern genetics

Brief Introduction of Evolutionary Algorithm

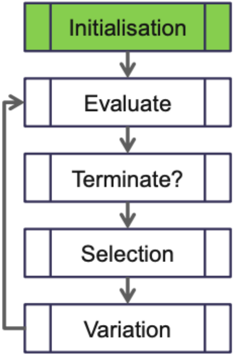
The following slides are from

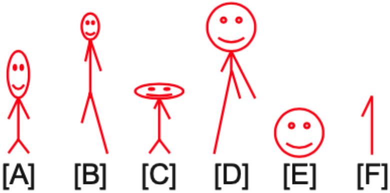
https://www.researchgate.net/publication/310365190_Introduction_to_Evolutionary_Algorithms





Brief Introduction of Evolutionary Algorithm





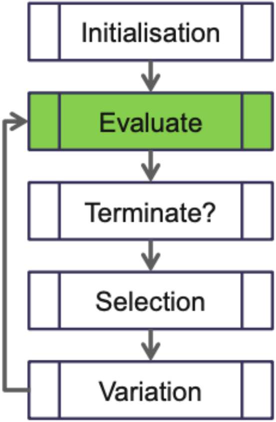
Parameters:

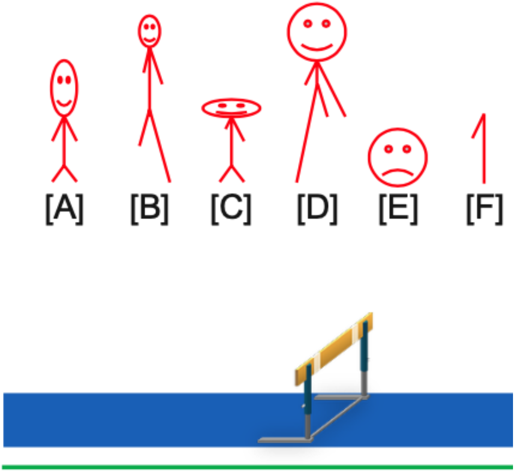
- Left leg length
- Right leg length
- Torso length
- Left arm length
- Right arm length
- Head Size

Chromosome:

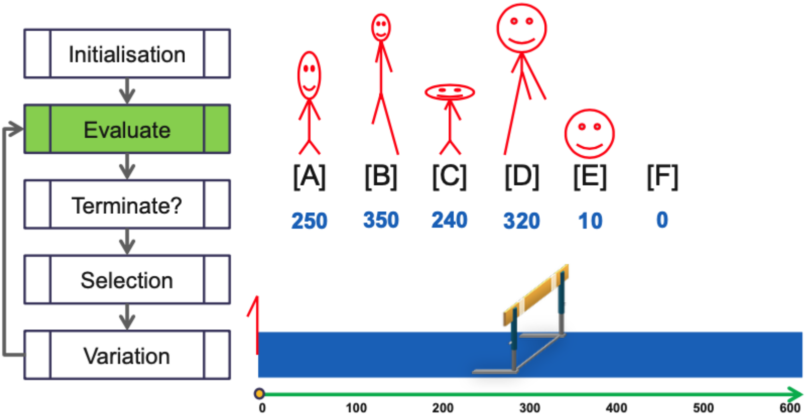
Left Leg	Torso	Left Arm	Head	Right Arm	Right Leg

Brief Introduction of Evolutionary Algorithm

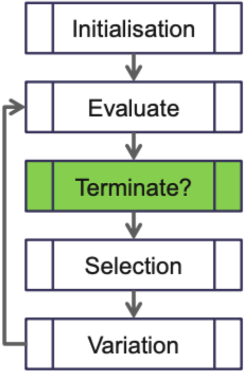


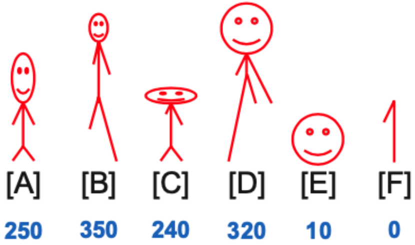


Brief Introduction of Evolutionary Algorithm



Brief Introduction of Evolutionary Algorithm

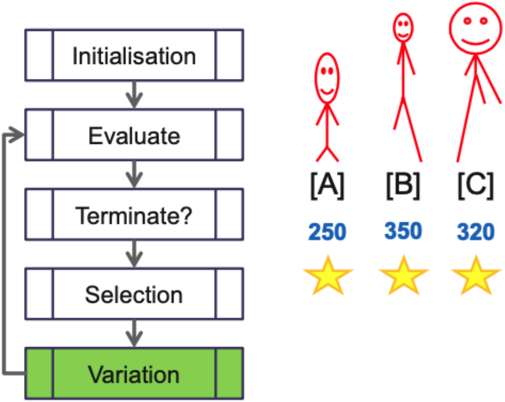




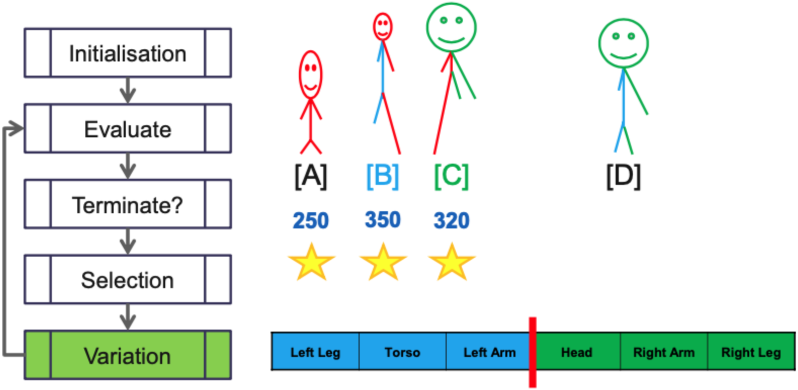
Termination Criteria

- Goal achieved?
- Number of generations reached max?
- Performance stagnating?

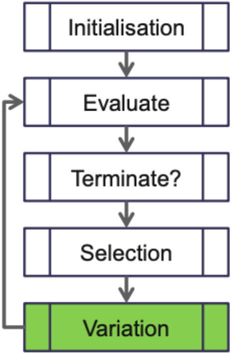
Brief Introduction of Evolutionary Algorithm

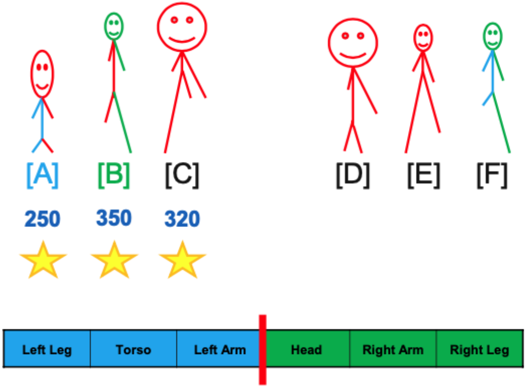


Brief Introduction of Evolutionary Algorithm

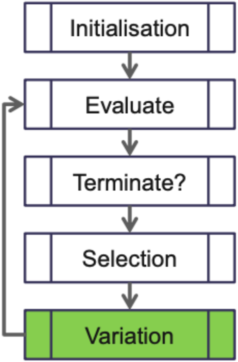


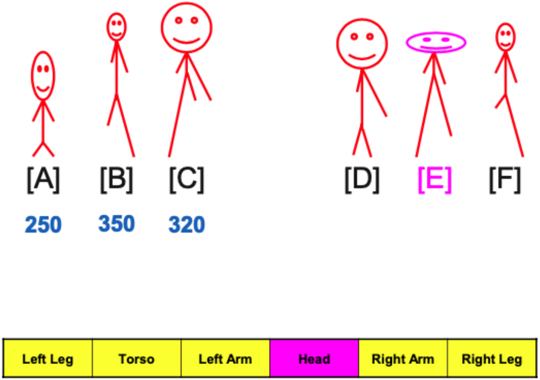
Brief Introduction of Evolutionary Algorithm



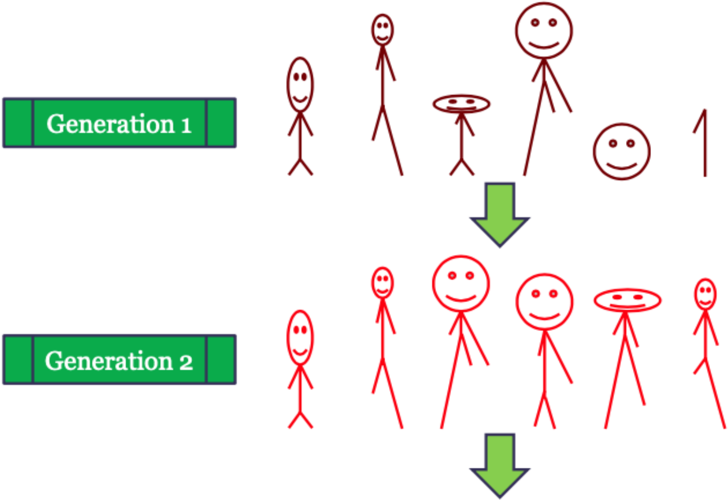


Brief Introduction of Evolutionary Algorithm





Brief Introduction of Evolutionary Algorithm



NAS with EA

Real et al., Regularized Evolution for Image Classifier Architecture Search. AAAI, 2019.

Steps:

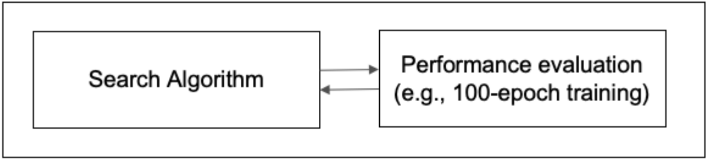
- Sample S models from the population
- Pick the one with the best performance as parent
- Mutate to generate child
- Train child, evaluate, add back to the population
- Discard the oldest from the population

Algorithm 1 Aging Evolution

```
population ← empty queue                          ▷ The population.
history ← ∅                                       ▷ Will contain all models.
while |population| < P do                         ▷ Initialize population.
    model.arch ← RANDOMARCHITECTURE()
    model.accuracy ← TRAINANDEVAL(model.arch)
    add model to right of population
    add model to history
end while
while |history| < C do                            ▷ Evolve for C cycles.
    sample ← ∅                                    ▷ Parent candidates.
    while |sample| < S do
        candidate ← random element from population
                              ▷ The element stays in the population.
        add candidate to sample
    end while
    parent ← highest-accuracy model in sample
    child.arch ← MUTATE(parent.arch)
    child.accuracy ← TRAINANDEVAL(child.arch)
    add child to right of population
    add child to history
    remove dead from left of population           ▷ Oldest.
    discard dead
end while
return highest-accuracy model in history
```

NAS is Expensive

- Experiments for NAS are typically time consuming to run
- RL or evolutionary algorithm often need to evaluate $>$ 10,000 configs in a single run



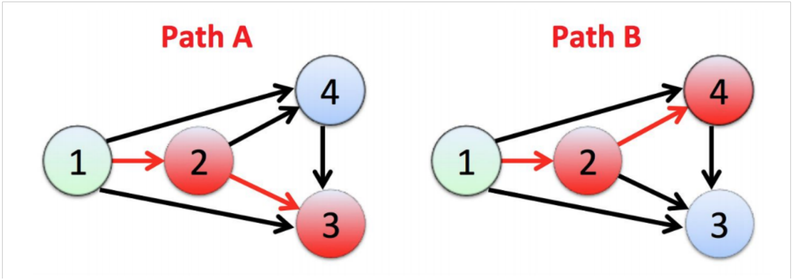
Differentiable NAS

- Significantly reduced search time since 2018

Architecture	Test Error (%)	Search Cost (GPU days)	Search Method
DenseNet-BC (Huang et al., 2017)	3.46	-	manual
NAS-RL (Zoph & Le, 2017)	3.65	22,400	RL
NASNet-A (Zoph et al., 2018)	2.65	2000	RL
BlockQNN (Zhong et al., 2018)	3.54	96	RL
AmoebaNet (Real et al., 2019)	3.34 ± 0.06	3150	evolution
Hierarchical GA (Liu et al., 2018)	3.75	300	evolution
GCP (Suganuma et al., 2017)	5.98	15	evolution
DARTS (1st) (Liu et al., 2019)	3.00 ± 0.14	0.4	differentiable
DARTS (2nd) (Liu et al., 2019)	2.76 ± 0.09	1.0	differentiable
SNAS (moderate) (Xie et al., 2019)	2.85 ± 0.02	1.5	differentiable
GDAS (Dong & Yang, 2019)	2.93	0.3	differentiable
ProxylessNAS (Cai et al., 2019)†	2.08	4.0	differentiable
PC-DARTS (Xu et al., 2020)	2.57 ± 0.07	0.1	differentiable
NASP (Yao et al., 2019)	2.83 ± 0.09	0.1	differentiable
SDARTS-ADV (Chen & Hsieh, 2020)	2.61 ± 0.02	1.3	differentiable
DrNAS (Chen et al., 2019)	2.46 ± 0.03	$0.6^\ddagger$	differentiable
DARTS+PT (Wang et al., 2020)	2.61 ± 0.08	0.8	differentiable

Can run on a single GPU machine!

Weight Sharing



- Models defined by Path A and Path B should be trained separately
- Can we assume Path A and Path B share the same weight at $1 \rightarrow 2$?

  Weight Sharing

  Avoid retraining for each new architecture

Conclusions

- A brief introduction to NAS
- Multiple search strategies

Questions?