Hierarchical Representations for Efficient Architecture Search

25/10/2018

Authors: Liu et al., 2018.

@ Carnegie Mellon University, Google Brain.

Presented by: Juan Carrillo

STAT946 Deep Learning

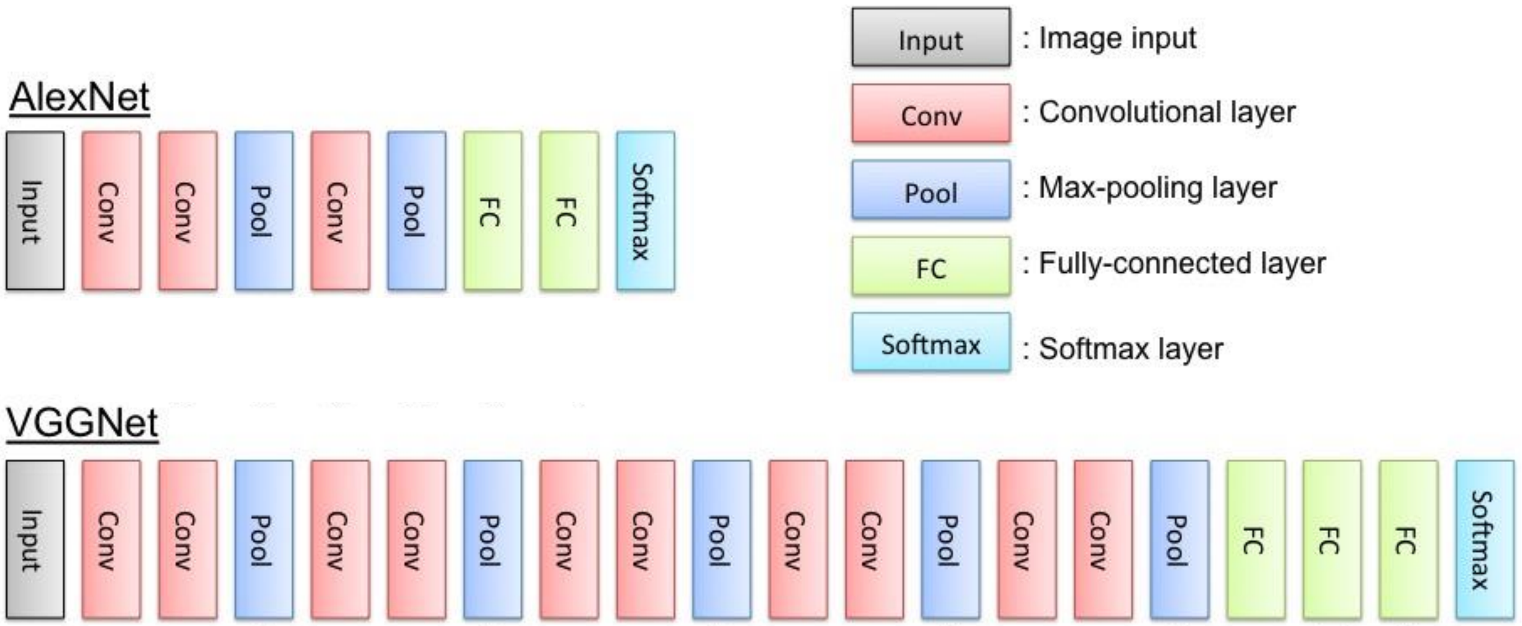
UNIVERSITY OF WATERLOO

Outline

- Introduction
- Architecture representations
- Evolutionary architecture search
- Experiments and results
- Contributions and critique

Introduction

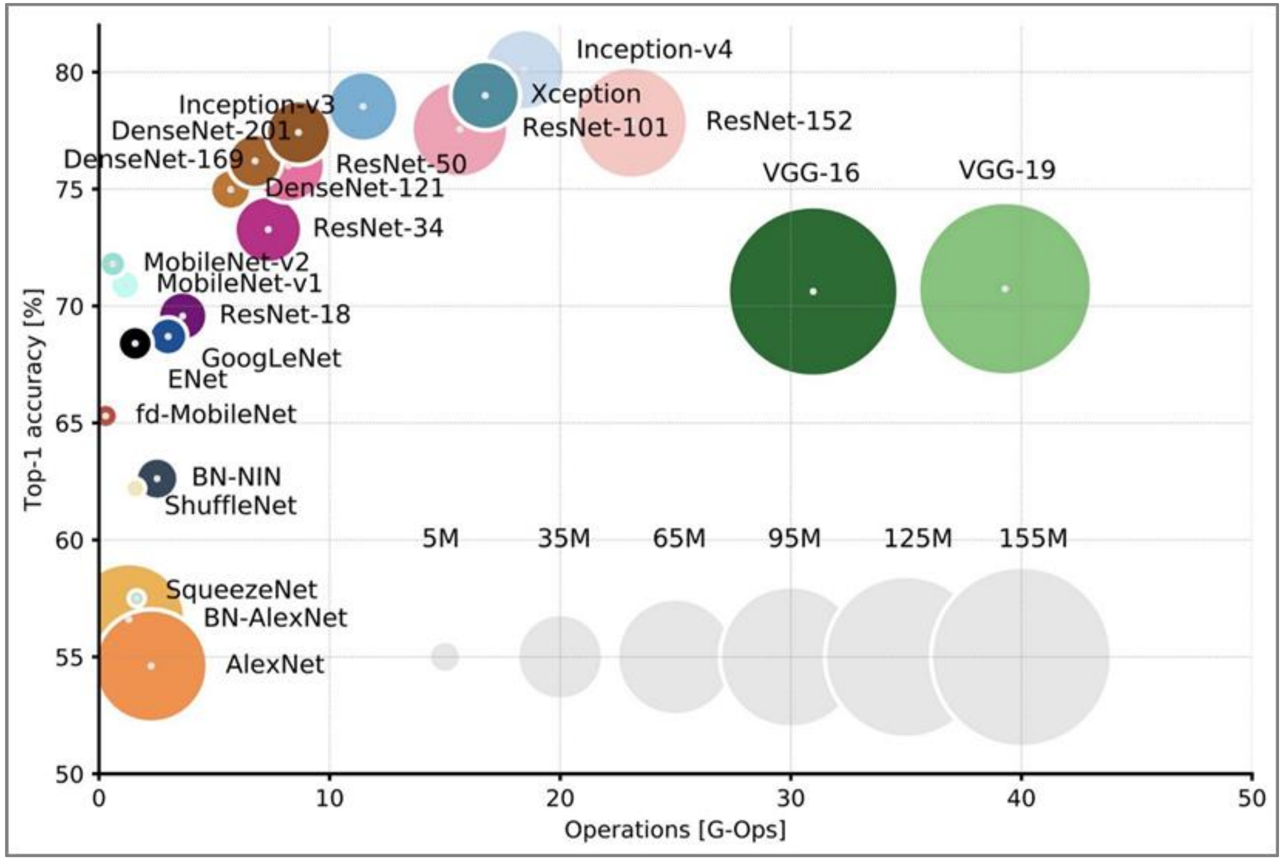
Deep Neural Network Architectures for Image Classification



Source: http://www.hirokatsukataoka.net/research/cnnfeatureevaluation/cnnfeatureevaluation.html

Introduction

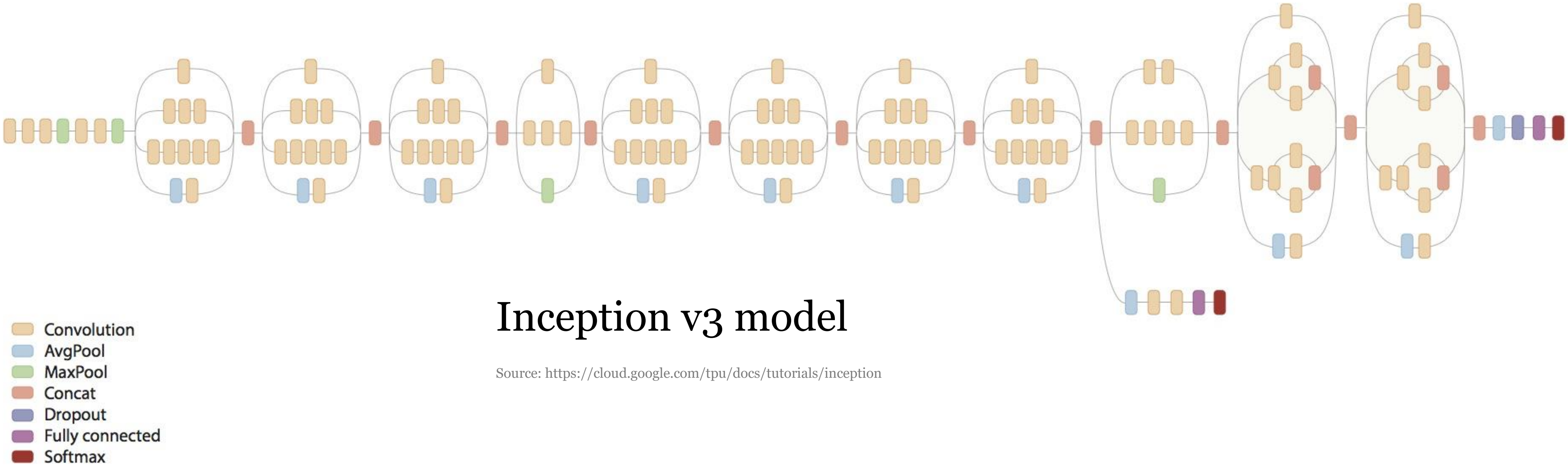
Computation cost versus accuracy



Source: https://arxiv.org/abs/1605.07678

Introduction

Increasing complexity of Deep Neural Networks



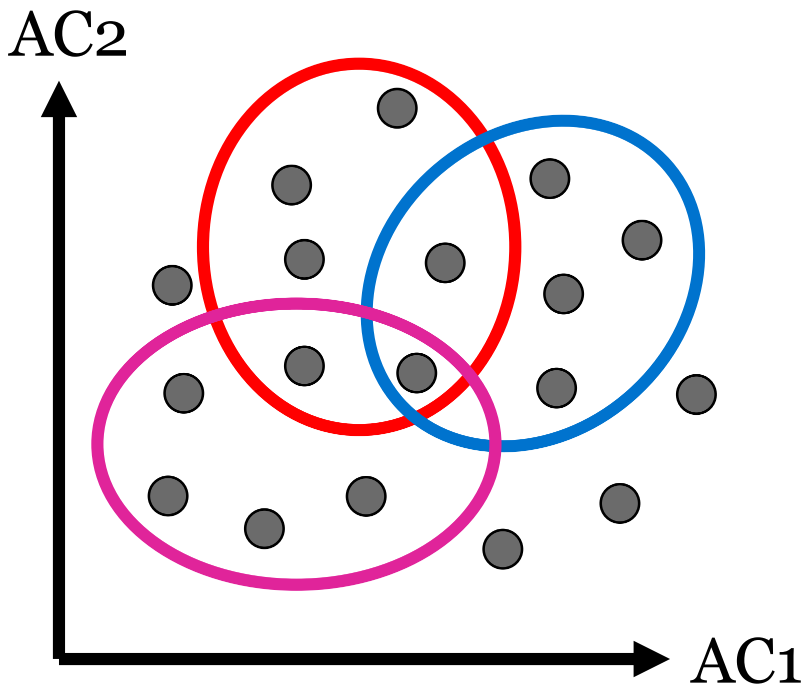
Inception v3 model

Source: https://cloud.google.com/tpu/docs/tutorials/inception

Introduction

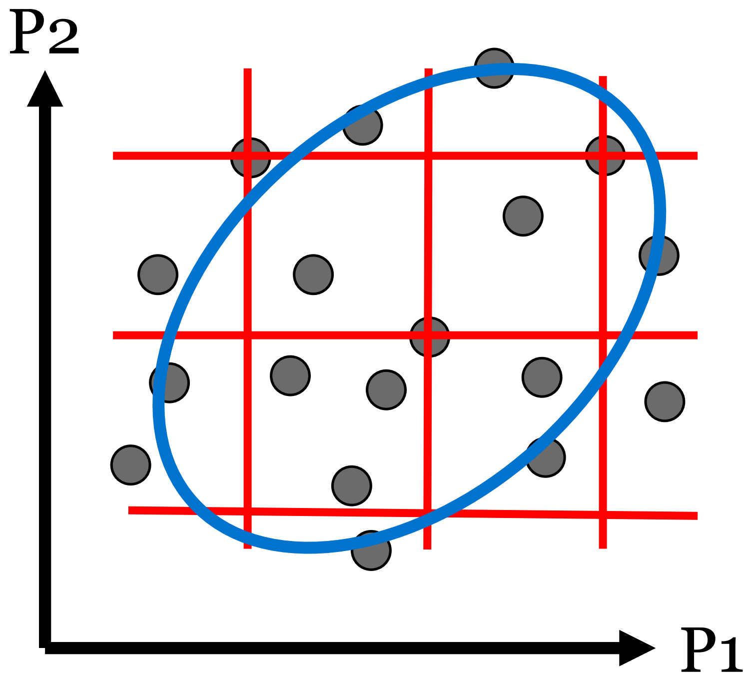
Architecture search

- MCTS
- RL
- Evolut.



Parameter search a.k.a. "Hyperparameter tuning"

- Grid
- Rand.



Architecture representations

Primitive operations

$$o^{(level)}_{operation\ id}$$

$o^{(1)}_1$ *1 x 1 convolution of C channels*

$o^{(2)}_2$ *3 x 3 depth wise convolution*

$o^{(1)}_3$ *3 x 3 separable convolution of C channels*

$o^{(1)}_4$ *3 x 3 max pooling*

$o^{(1)}_5$ *3 x 3 average pooling*

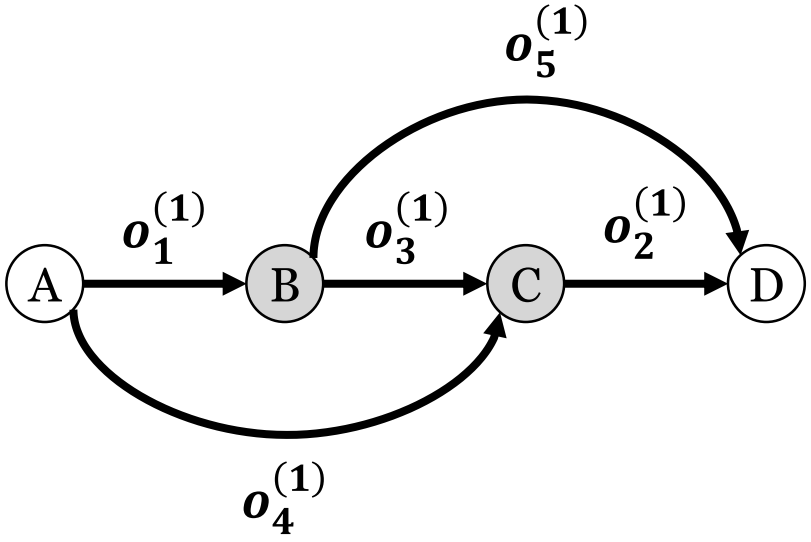
$o^{(1)}_6$ *identity*

Convolutional operations are followed by batch normalization and ReLU activation

Architecture representations

Flat representation

- Nodes are feature maps (a.k.a. vectors or matrices)
- Edges are primitive operations



To...

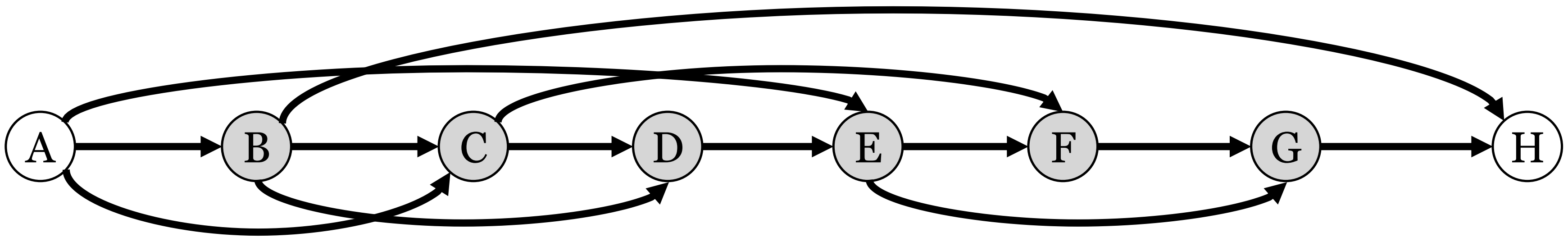
From...

$$
\begin{array}{c c}
 & \begin{array}{ccc} B & C & D \end{array} \\
\begin{array}{c} A \\ B \\ C \end{array} & \begin{bmatrix} \boldsymbol{o}_1^{(1)} & \boldsymbol{o}_4^{(1)} & . \\ . & \boldsymbol{o}_3^{(1)} & \boldsymbol{o}_5^{(1)} \\ . & . & \boldsymbol{o}_2^{(1)} \end{bmatrix}
\end{array}
$$

Architecture representations

Flat representation

- Flat networks can grow to very complex models.



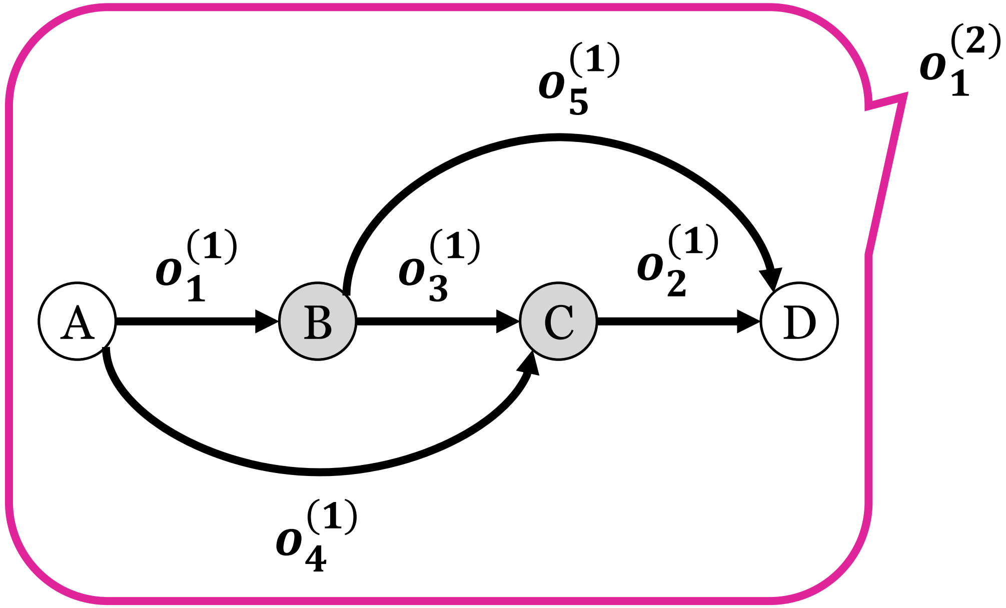
UNIVERSITY OF WATERLOO

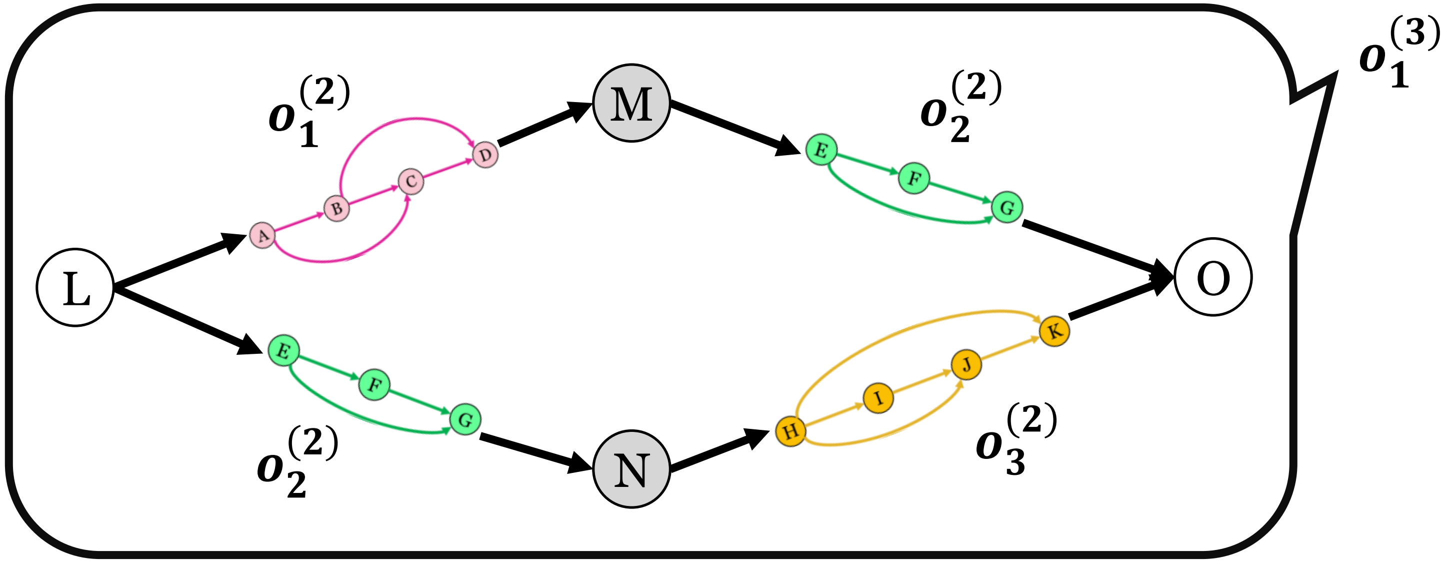
Architecture representations

Hierarchical representation

- Nodes are feature maps (a.k.a. vectors or matrices)
- Edges are higher level operations (a.k.a. motifs)

$$\begin{array}{c} \\ L \\ M \\ N \end{array} \begin{array}{c} \begin{array}{ccc} M & N & O \end{array} \\ \begin{bmatrix} \boldsymbol{o}_1^{(2)} & \boldsymbol{o}_2^{(2)} & . \\ . & . & \boldsymbol{o}_2^{(2)} \\ . & . & \boldsymbol{o}_3^{(2)} \end{bmatrix} \end{array}$$

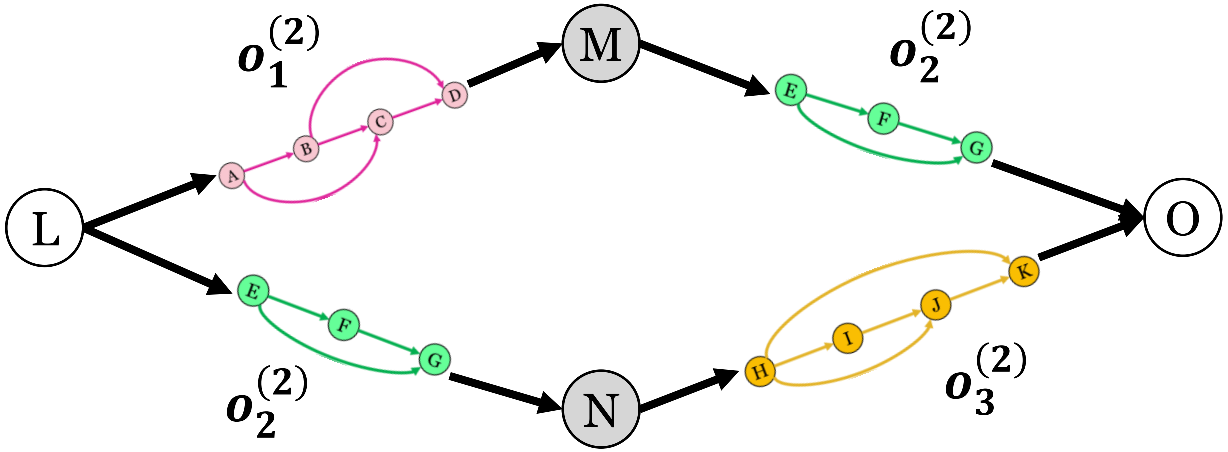




Evolutionary architecture search

Mutation

- Any NN architecture is treated as a genotype in a population.



Sample a non-primitive level

↓

Sample a motif

↓

Sample two nodes

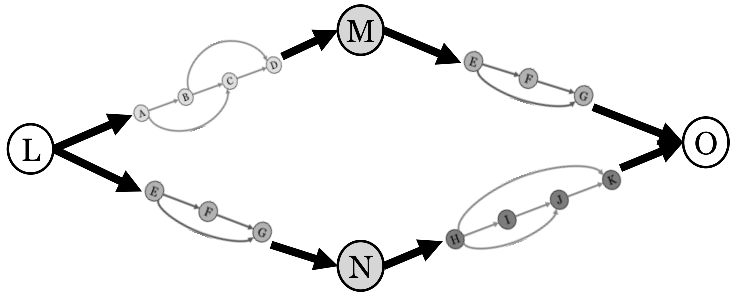
↓

Change current operation

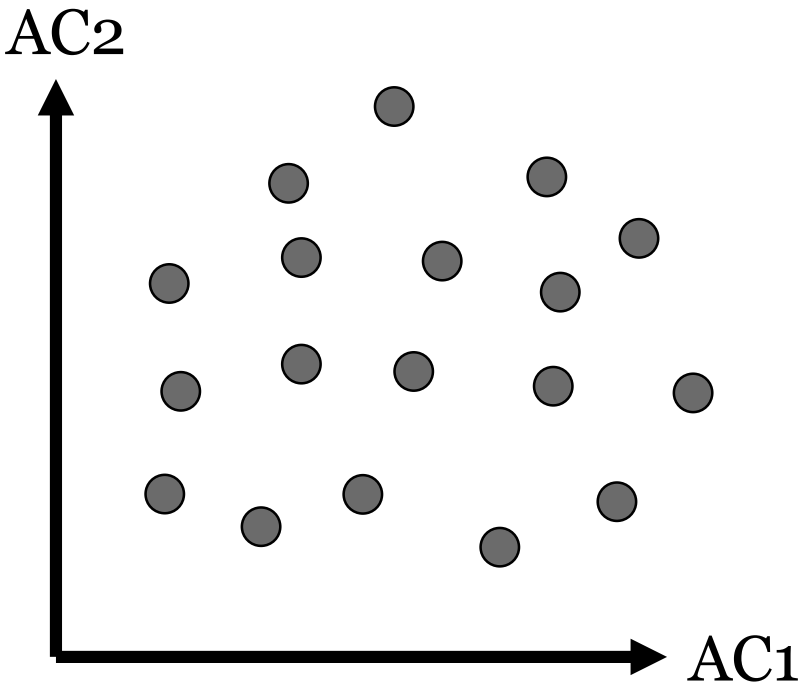
Evolutionary architecture search

Initialization

Trivial genotype where all operations are identity



Mutation

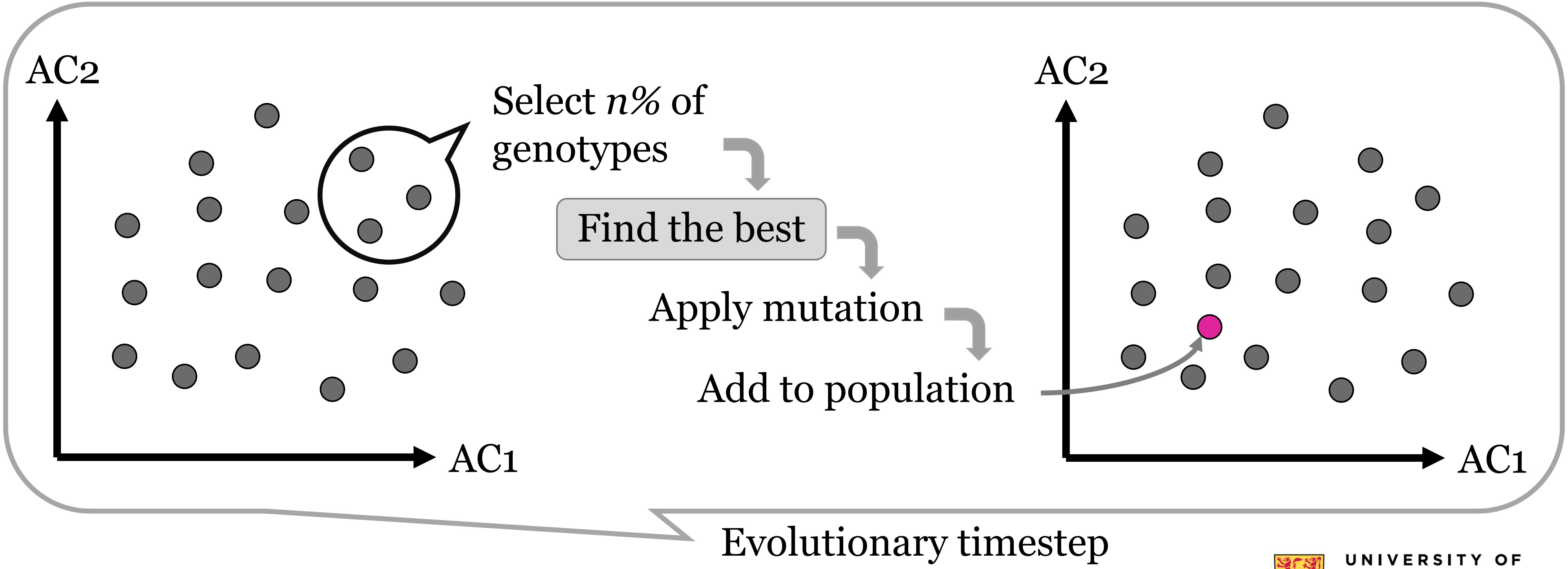


Initialization timestep

UNIVERSITY OF WATERLOO

Evolutionary architecture search

Search algorithm: Tournament selection



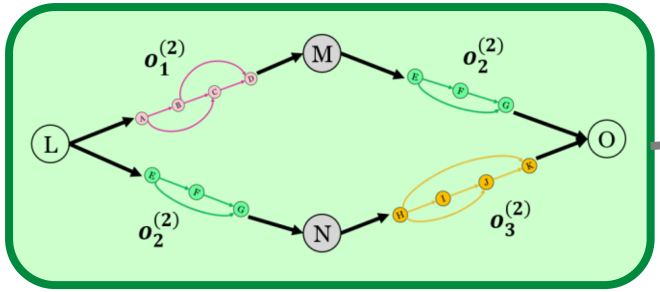
Experiments and results

Architecture search on small CIFAR-10

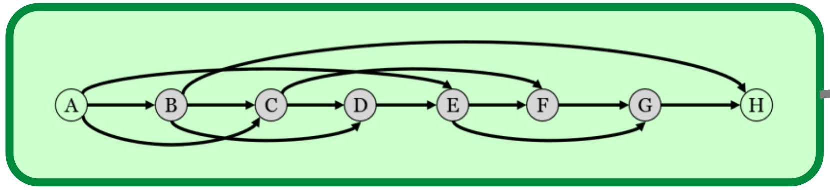
Host model: Small CIFAR-10

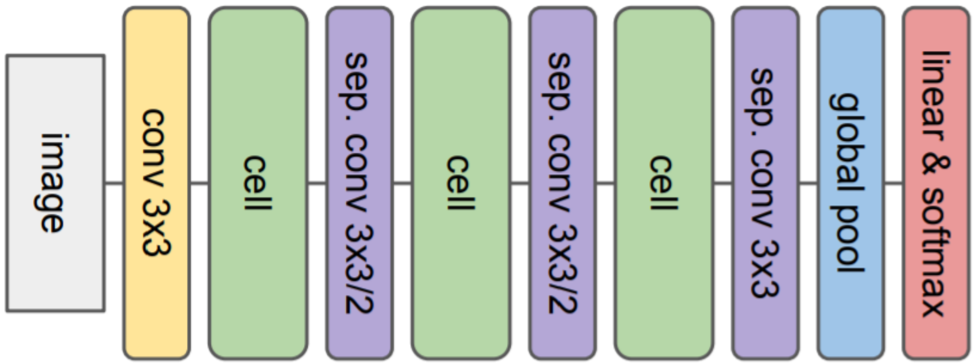
Cell

Hierarchical rep.



Flat rep.





5000 training steps (SGD) and evaluate fitness.

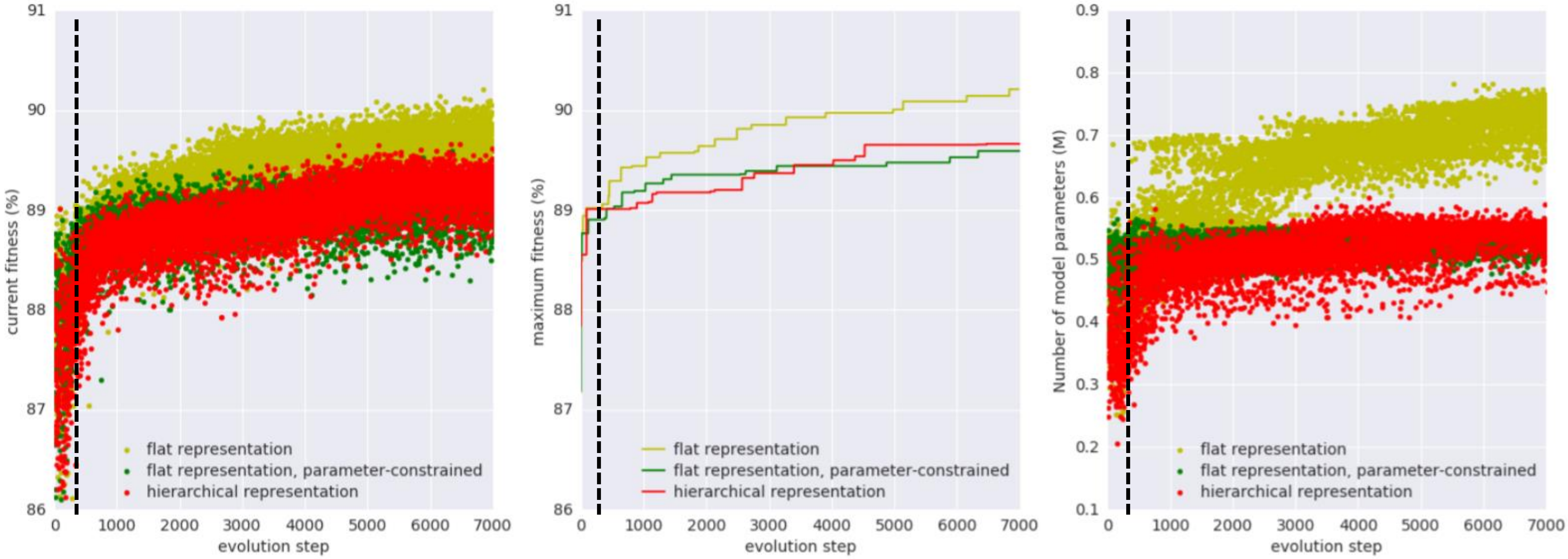
Average of four training-evaluation runs.

Experiments and results

Step 1 to step 200: Initialization

Step 201 to step 7000: Evolution

Architecture search on small CIFAR-10



Experiments and results

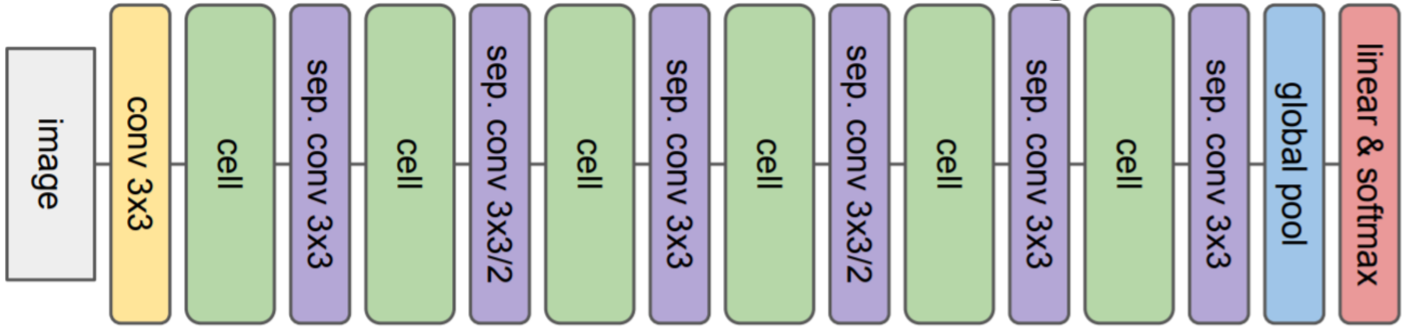
Architecture evaluation on large CIFAR-10 and ImageNet

Best Hierarchical Cell

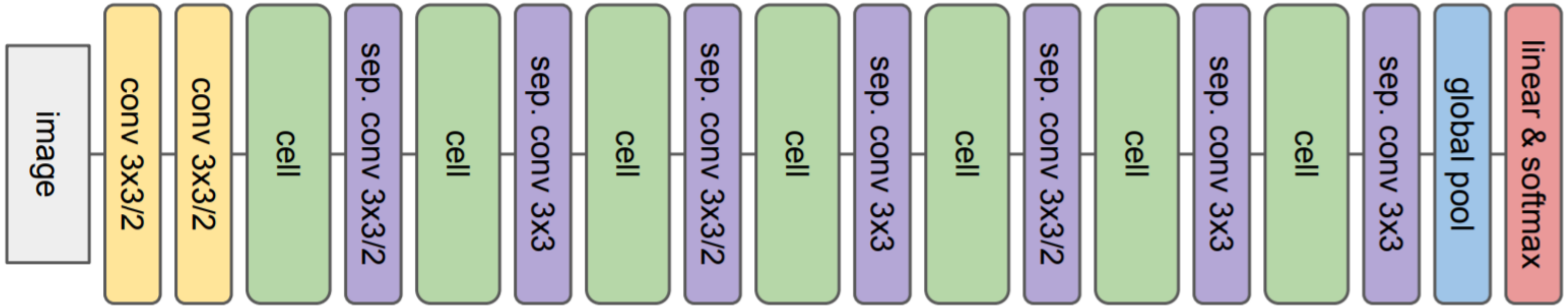
Best Flat Cell

Best Flat constr. Cell

Host model: Large CIFAR-10



Host model: ImageNet



Experiments and results

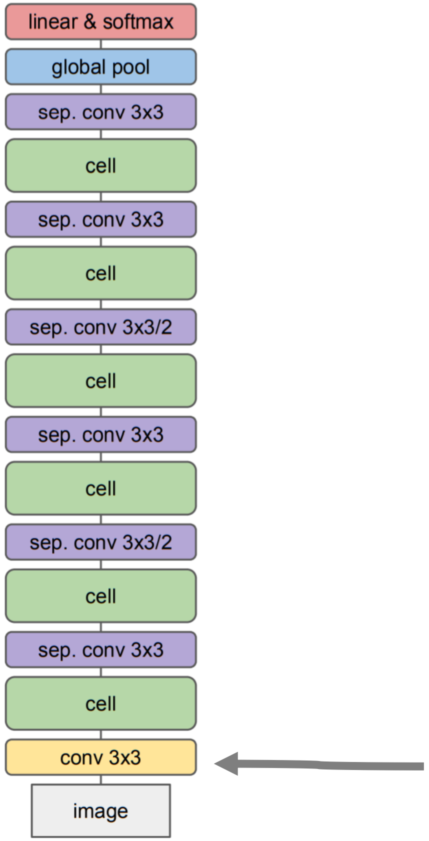
Architecture evaluation on large CIFAR-10 and ImageNet

Search Method	CIFAR-10 error (%)	ImageNet Top-1/Top-5 error (%)
Flat repr-n, random architecture	4.56 ± 0.11	21.4/5.8
Flat repr-n, random search (200 samples)	4.02 ± 0.11	20.8/5.7
Flat repr-n, evolution (7000 samples)	3.92 ± 0.06	20.6/5.6
Flat repr-n, parameter-constrained, evolution (7000 samples)	4.17 ± 0.08	21.2/5.8
Hier. repr-n, random architecture	4.21 ± 0.11	21.5/5.8
Hier. repr-n, random search (200 samples)	4.04 ± 0.2	20.4/5.3
Hier. repr-n, random search (7000 samples)	3.91 ± 0.15	21.0/5.5
Hier. repr-n, evolution (7000 samples)	$\mathbf{3.75 \pm 0.12}$	**20.3/5.2**

The best type of cells is?

Experiments and results

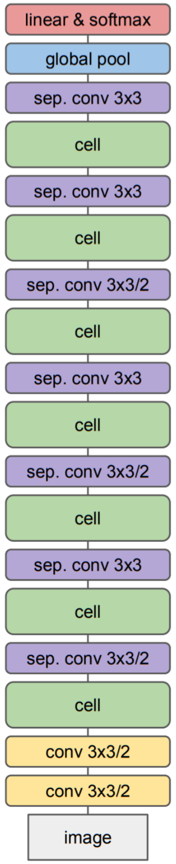
Comparison against other models on CIFAR-10



Model	Error (%)
ResNet-1001 + pre-activation (He et al., 2016b)	4.62
Wide ResNet-40-10 + dropout (Zagoruyko & Komodakis, 2016)	3.8
DenseNet (k=24) (Huang et al., 2016)	3.74
DenseNet-BC (k=40) (Huang et al., 2016)	3.46
MetaQNN (Baker et al., 2016)	6.92
NAS v3 (Zoph & Le, 2016)	3.65
Block-QNN-A (Zhong et al., 2017)	3.60
NASNet-A (Zoph et al., 2017)	3.41
Evolving DNN (Miikkulainen et al., 2017)	7.3
Genetic CNN (Xie & Yuille, 2017)	7.10
Large-scale Evolution (Real et al., 2017)	5.4
SMASH (Brock et al., 2017)	4.03
Evolutionary search, hier. repr., $c_0 = 64$	3.75 ± 0.12
Evolutionary search, hier. repr., $c_0 = 128$	3.63 ± 0.10

Experiments and results

Comparison against other models on ImageNet



Model	Top-1 error (%)	Top-5 error (%)
Inception-v3 (Szegedy et al., 2016)	21.2	5.6
Xception (Chollet, 2016)	21.0	5.5
Inception-ResNet-v2 (Szegedy et al., 2017)	19.9	4.9
NASNet-A (Zoph et al., 2017)	19.2	4.7
Evolutionary search, hier. repr., $c_0 = 64$	20.3	5.2

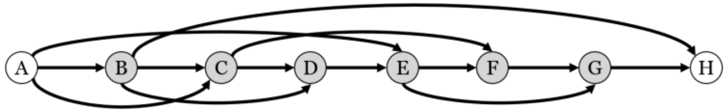
Contributions

A new evolutionary framework is introduced for searching neural network architectures over searching spaces defined by flat and hierarchical representations of a convolutional cell.

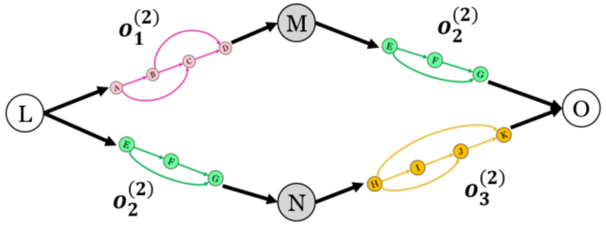
Experiments show that the proposed framework achieves competitive results against state-of-the-art classifiers on the CIFAR-10 and ImageNet datasets.

Critique

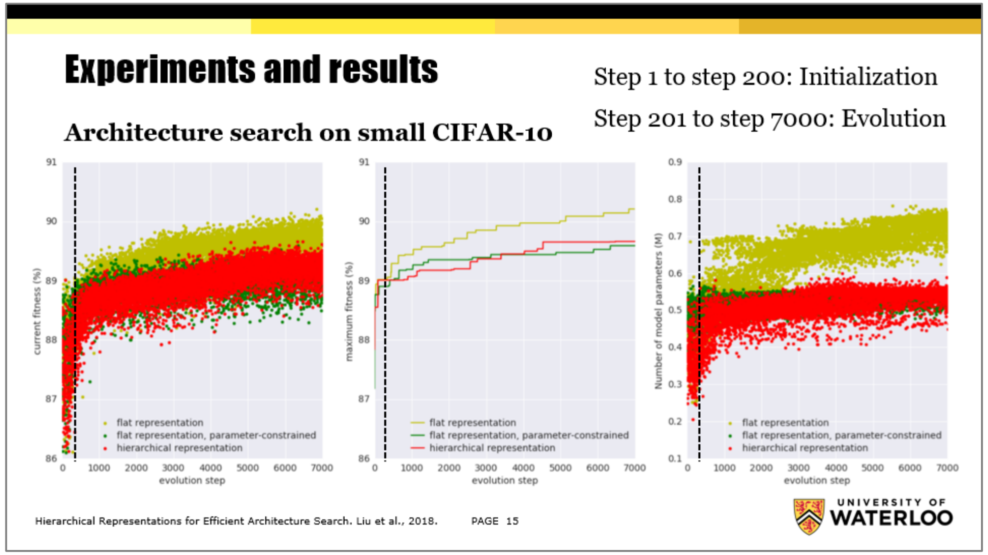
What is the best type of cell?



Flat representation



Hierarchical representation



Experiments and results

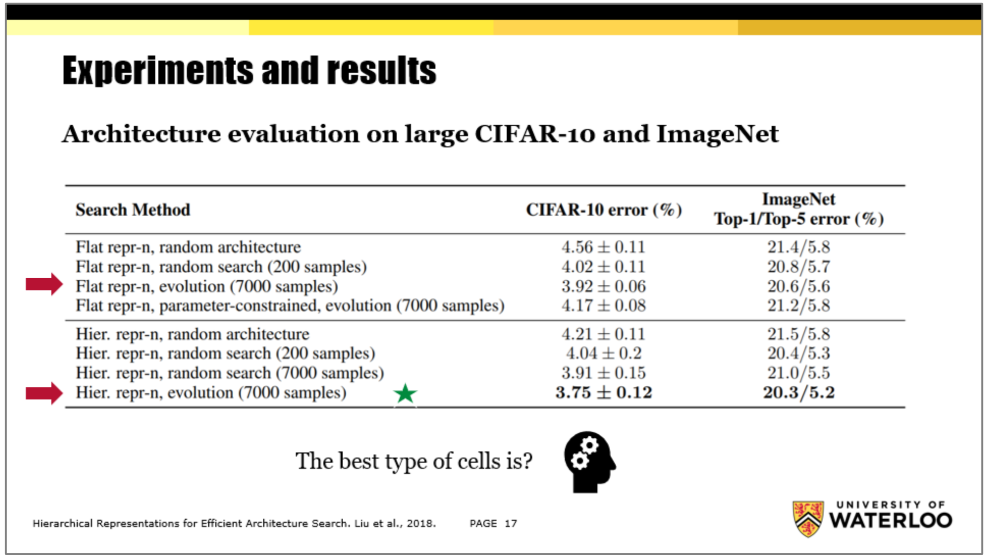
Step 1 to step 200: Initialization

Step 201 to step 7000: Evolution

Architecture search on small CIFAR-10

Hierarchical Representations for Efficient Architecture Search. Liu et al., 2018. PAGE 15

UNIVERSITY OF WATERLOO



Experiments and results

Architecture evaluation on large CIFAR-10 and ImageNet

Search Method	CIFAR-10 error (%)	ImageNet Top-1/Top-5 error (%)
Flat repr-n, random architecture	4.56 ± 0.11	21.4/5.8
Flat repr-n, random search (200 samples)	4.02 ± 0.11	20.8/5.7
Flat repr-n, evolution (7000 samples)	3.92 ± 0.06	20.6/5.6
Flat repr-n, parameter-constrained, evolution (7000 samples)	4.17 ± 0.08	21.2/5.8
Hier. repr-n, random architecture	4.21 ± 0.11	21.5/5.8
Hier. repr-n, random search (200 samples)	4.04 ± 0.2	20.4/5.3
Hier. repr-n, random search (7000 samples)	3.91 ± 0.15	21.0/5.5
Hier. repr-n, evolution (7000 samples)	$\mathbf{3.75 \pm 0.12}$	**20.3/5.2**

The best type of cells is?

Hierarchical Representations for Efficient Architecture Search. Liu et al., 2018. PAGE 17

UNIVERSITY OF WATERLOO

UNIVERSITY OF WATERLOO

Questions?

Source: https://www.theacademygtc.co.uk/thinking-man-wasnt-a-thing/