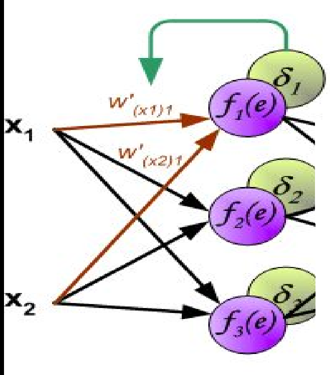
Instructors

	Fabio Gonzalez Full Professor UNIVERSIDAD NACIONAL DE COLOMBIA National University of Colombia Visiting Professor at UH Email: fagonzalezo@unal.edu.co Office: PGH 598
	Thamar Solorio UNIVERSITY of HOUSTON Associate Professor University of Houston Email: thamar.solorio@gmail.com Office: PGH 584

Intro to DL

Backpropagation



$$w'_{(x1)1} = w_{(x1)1} + \eta \delta_1 \frac{df_1(e)}{de} x_1$$

$$w'_{(x2)1} = w_{(x2)1} + \eta \delta_1 \frac{df_1(e)}{de} x_2$$

letters to nature

nature

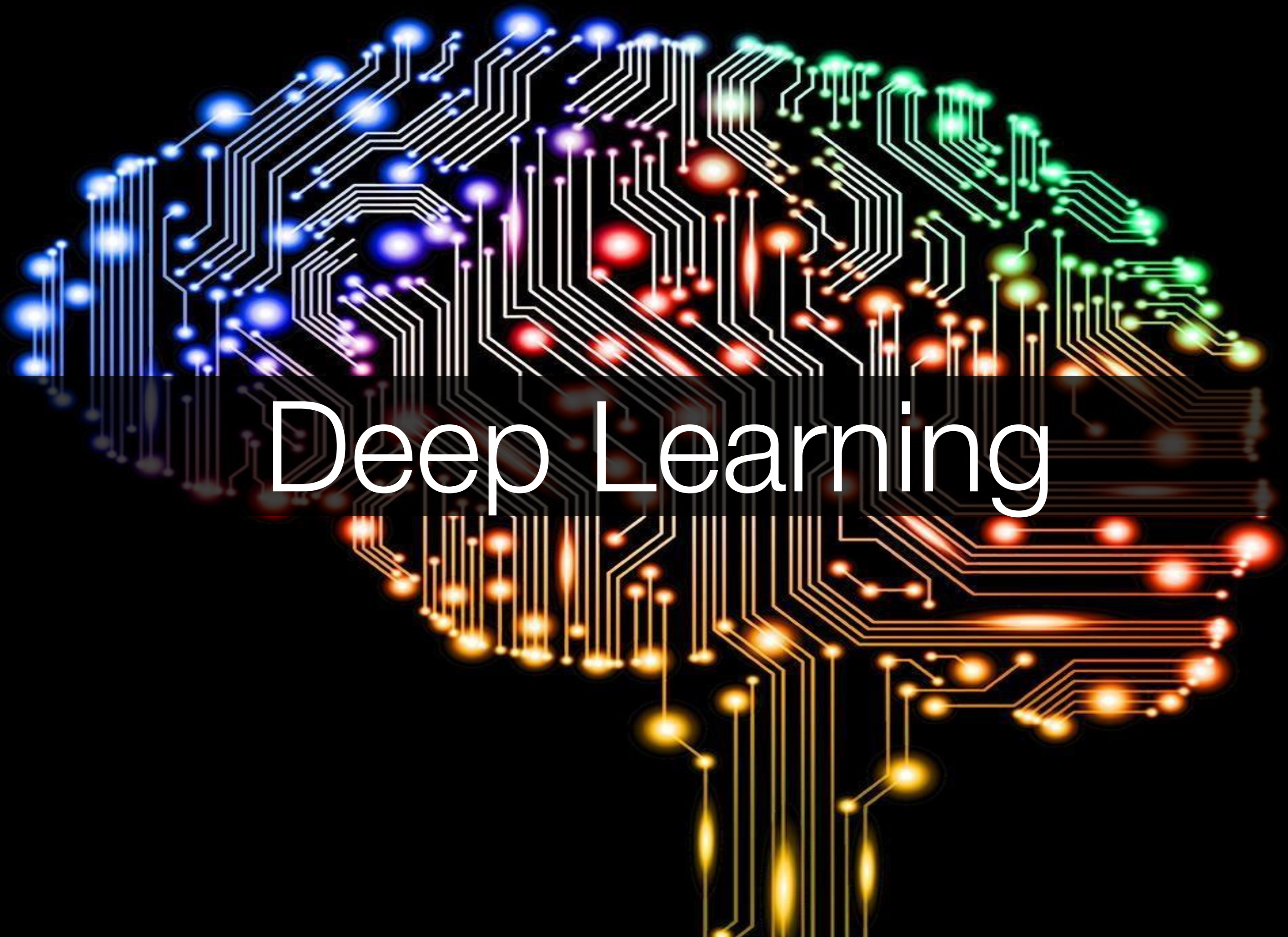
Nature **323**, 533 - 536 (09 October 1986); doi:10.1038/323533a0

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of Californi
San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon Universit
Pittsburgh, Philadelphia 15213, USA

Source: http://home.agh.edu.pl/~vlsi/AI/backp_t_en/backprop.html


Deep Learning

Deep learning boom

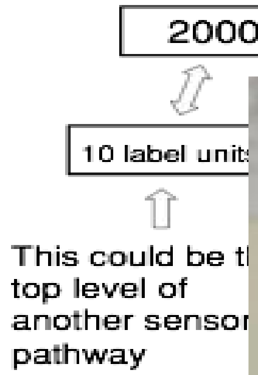
Deep Learning is Born

Neural Computation 18, 1527–1554 (2006) © 2006 Massachusetts Institute of Technology

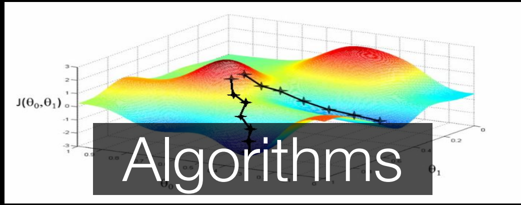
A Fast Learning Algori

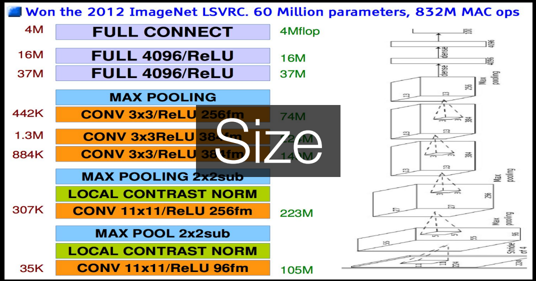
Geoffrey E. Hinton
hinton@cs.toronto.edu
Simon Osindero
osindero@cs.toronto.edu
Department of Computer Scien

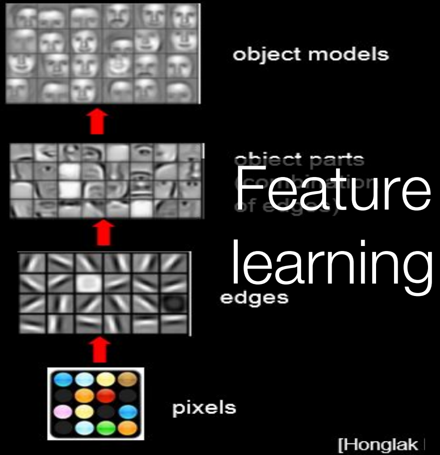
Yee-Whye Teh
tehyw@comp.nus.edu.sg
Department of Computer Scien

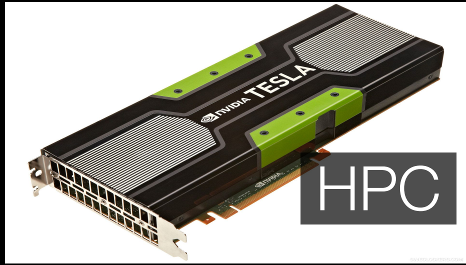


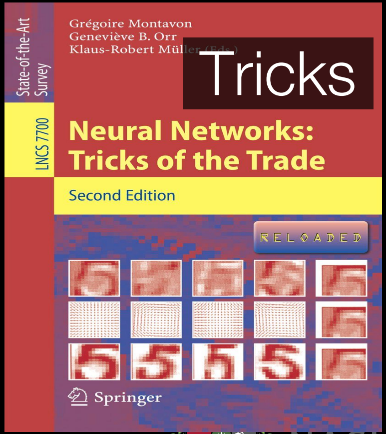
Deep learning recipe



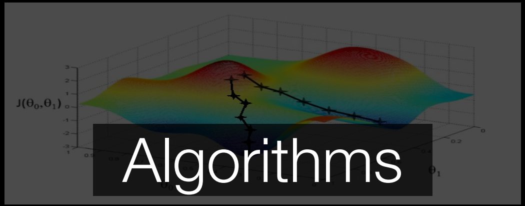


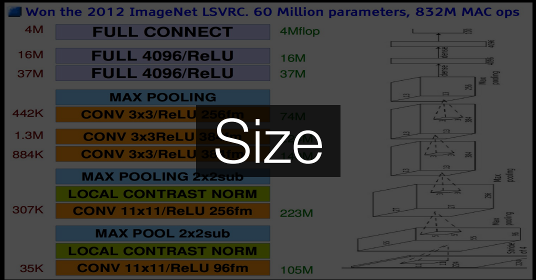


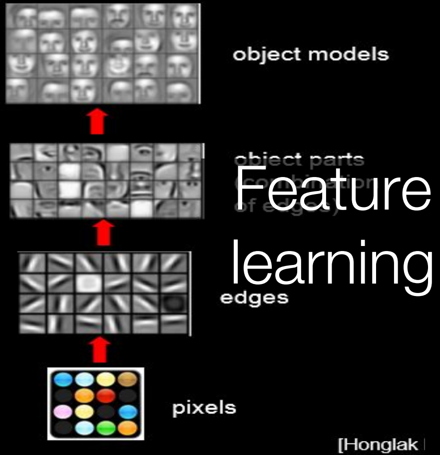


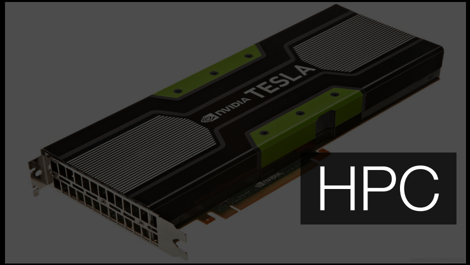


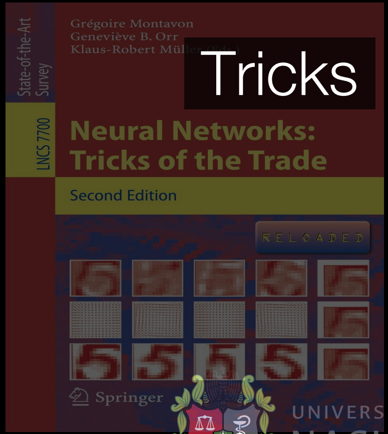
Deep learning recipe







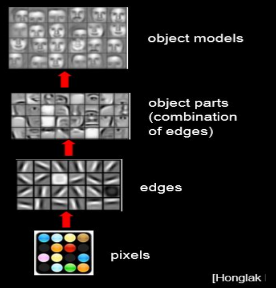


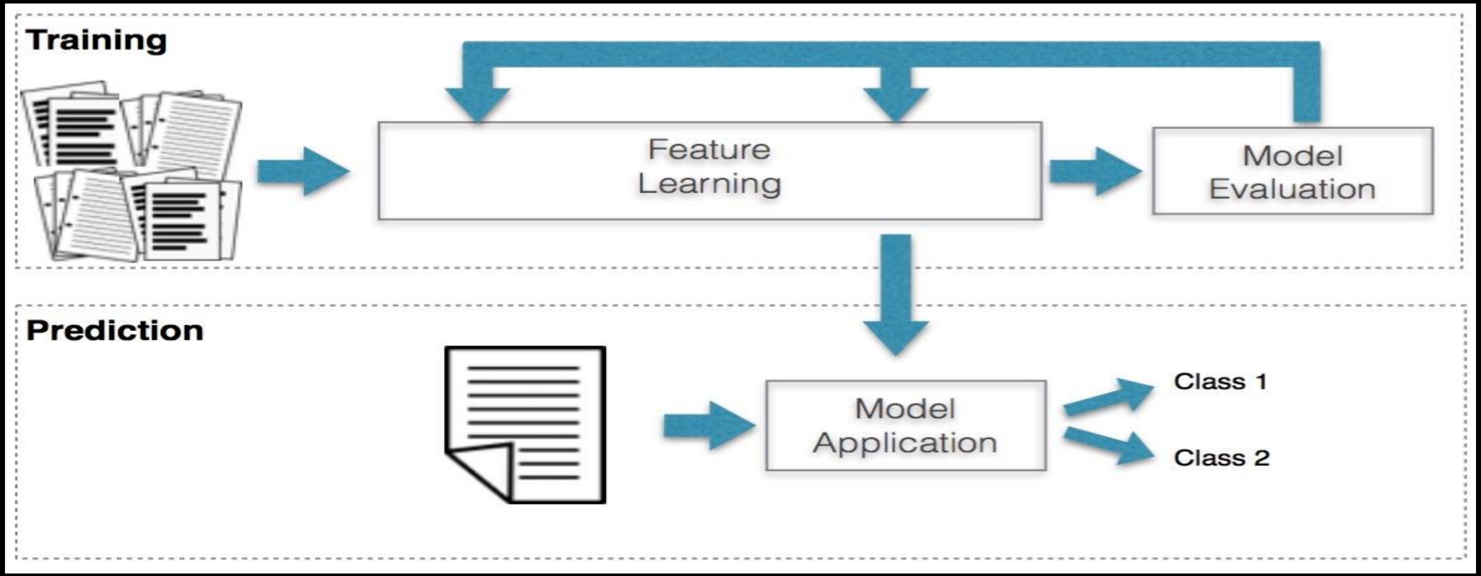


UNIV HOUSTON

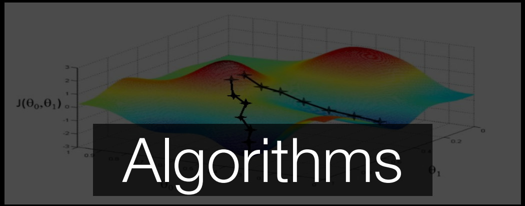
UNIVERSIDAD NACIONAL DE COLOMBIA

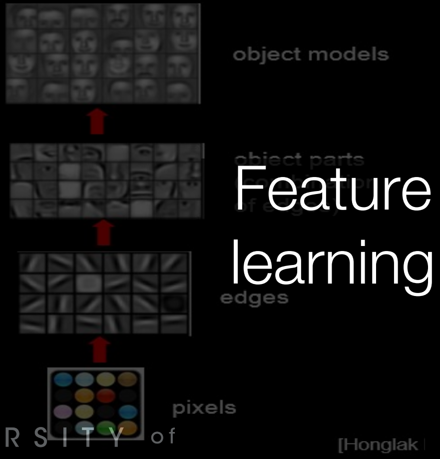
Feature learning

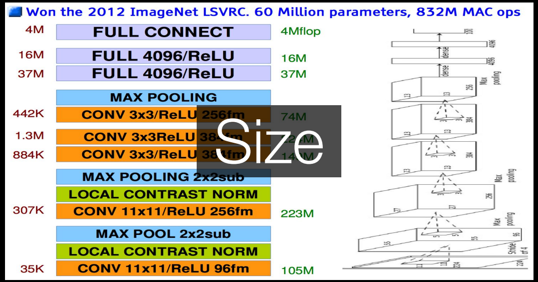


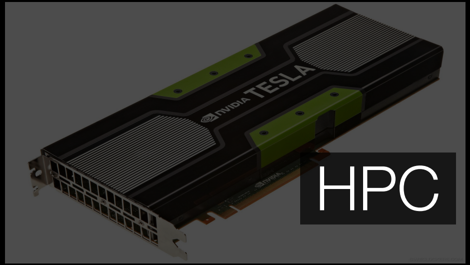


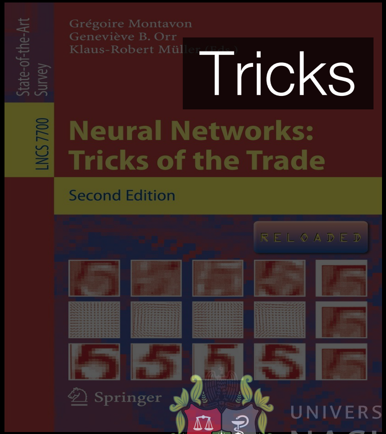
Deep learning recipe







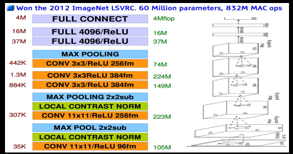


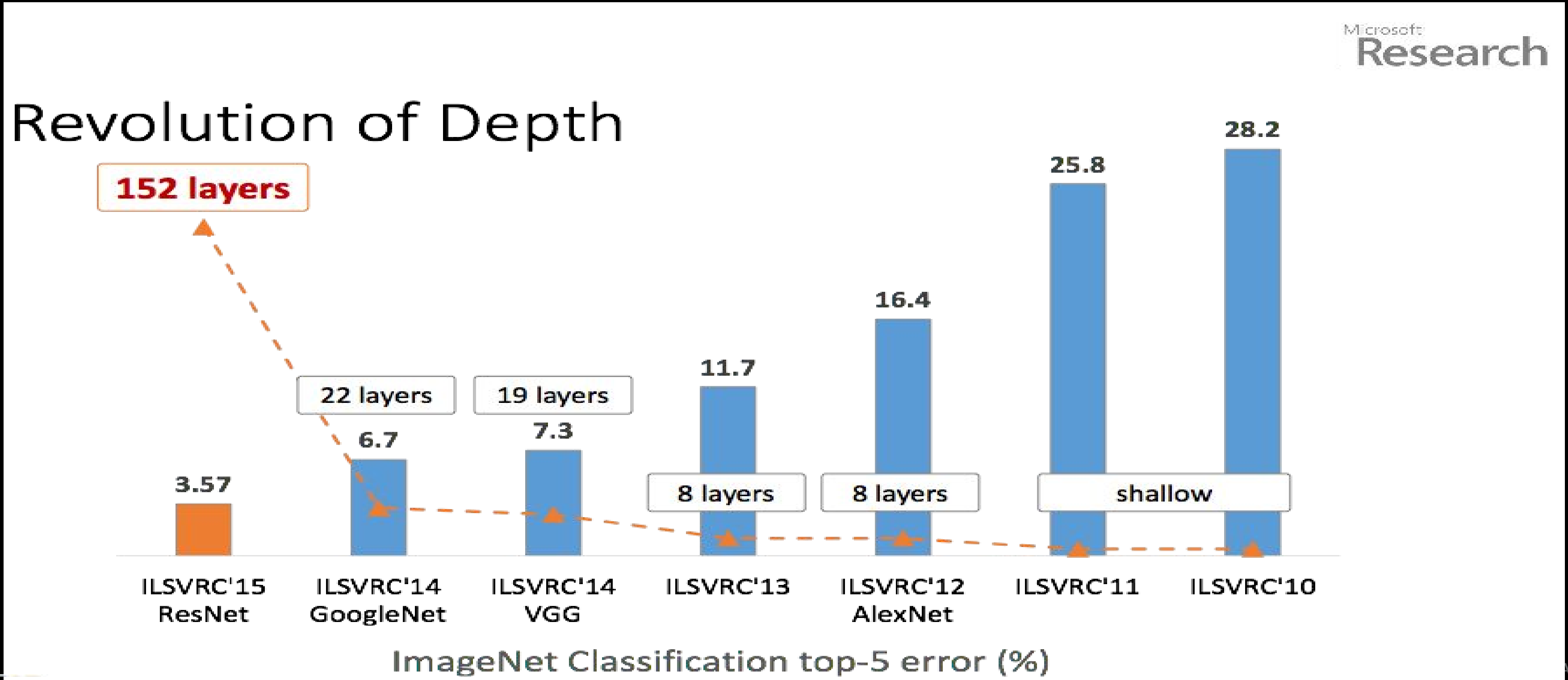


UNIVERSITY of HOUSTON

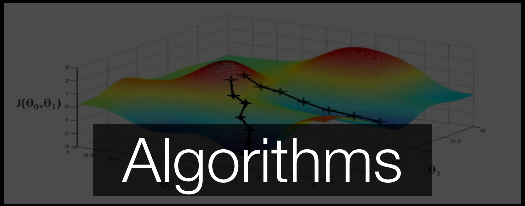
UNIVERSIDAD NACIONAL DE COLOMBIA

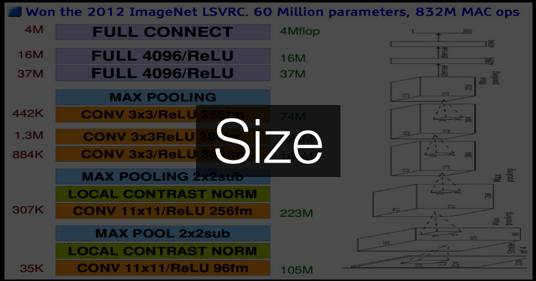
Deep —> Bigger

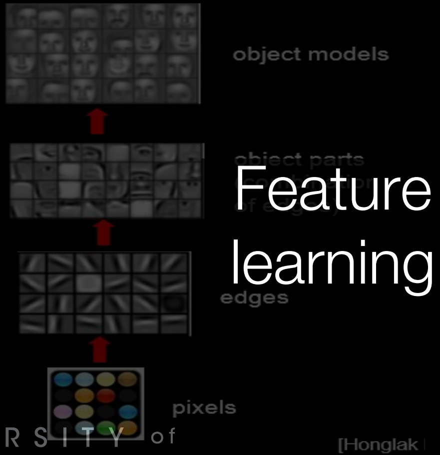

Won the 2012 ImageNet LSVRC. 60 Million parameters, 832M MAC ops
4M
FULL CONNECT
4Mflop
16M
FULL 4096/ReLU
16M
37M
FULL 4096/ReLU
37M
MAX POOLING
442K
CONV 3x3/ReLU 256fm
74M
1.3M
CONV 3x3ReLU 384fm
224M
884K
CONV 3x3/ReLU 384fm
149M
MAX POOLING 2x2sub
LOCAL CONTRAST NORM
307K
CONV 11x11/ReLU 256fm
223M
MAX POOL 2x2sub
LOCAL CONTRAST NORM
35K
CONV 11x11/ReLU 96fm
105M

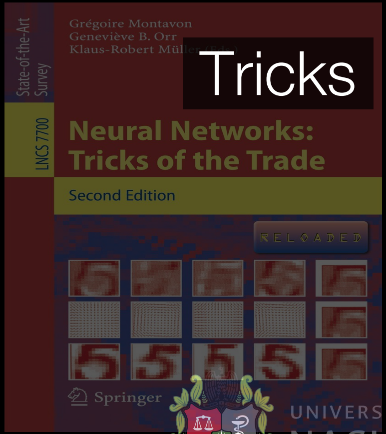

Microsoft
Research
Revolution of Depth
152 layers
22 layers
19 layers
8 layers
8 layers
shallow
3.57
6.7
7.3
11.7
16.4
25.8
28.2
ILSVRC'15
ResNet
ILSVRC'14
GoogleNet
ILSVRC'14
VGG
ILSVRC'13
ILSVRC'12
AlexNet
ILSVRC'11
ILSVRC'10
ImageNet Classification top-5 error (%)

Deep learning recipe







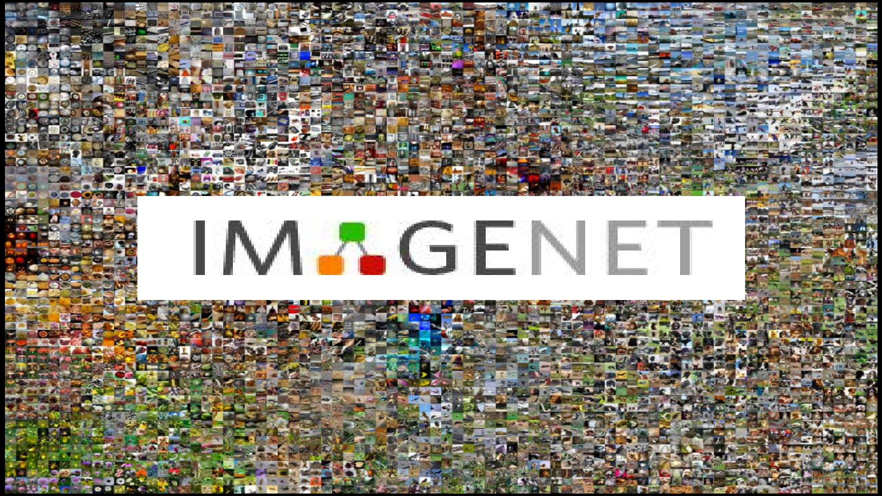


UNIVERSITY of HOUSTON

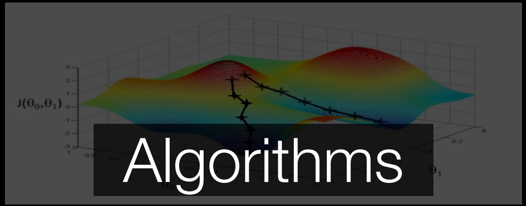
UNIVERSIDAD NACIONAL DE COLOMBIA

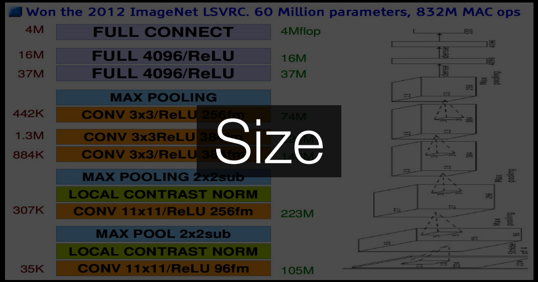
Data...

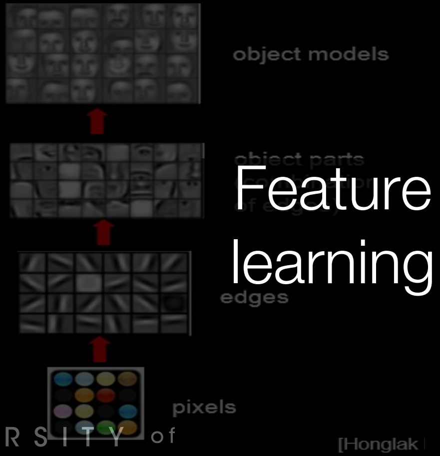
- Images annotated with WordNet concepts
- Concepts: 21,841
- Images: 14,197,122
- Bounding box annotations: 1,034,908
- Crowdsourcing

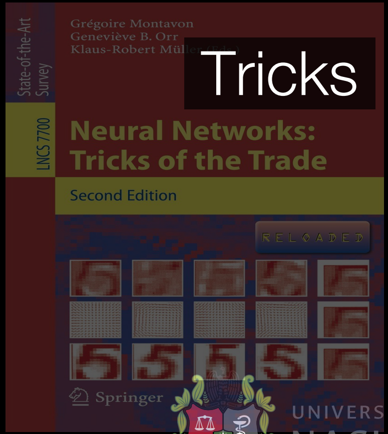


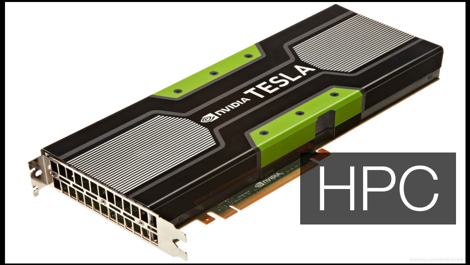
Deep learning recipe



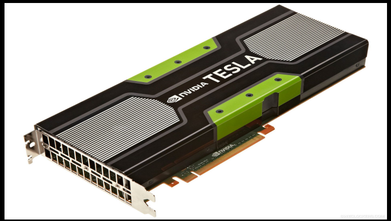








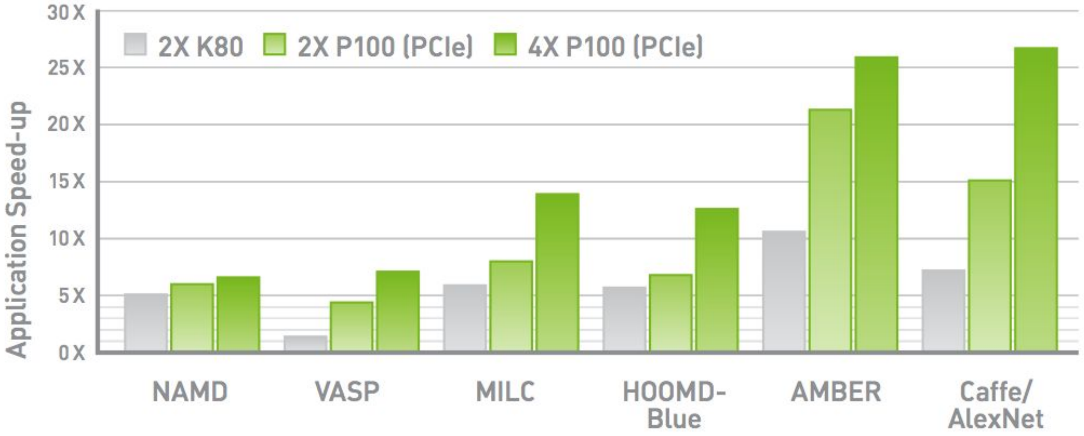
HPC



SPECIFICATIONS

GPU Architecture	**NVIDIA Pascal**
NVIDIA CUDA® Cores	**3584**
Double-Precision Performance	**4.7 TeraFLOPS**
Single-Precision Performance	**9.3 TeraFLOPS**
Half-Precision Performance	**18.7 TeraFLOPS**
GPU Memory	**16GB CoWoS HBM2 at 732 GB/s or 12GB CoWoS HBM2 at 549 GB/s**
System Interface	**PCIe Gen3**
Max Power Consumption	**250 W**
ECC	**Yes**
Thermal Solution	**Passive**
Form Factor	**PCIe Full Height/Length**

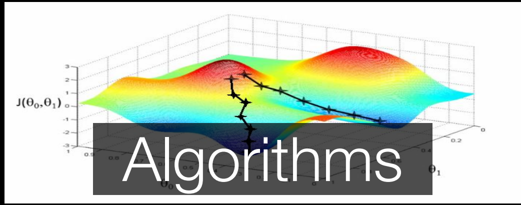
NVIDIA Tesla P100 for PCIe Performance

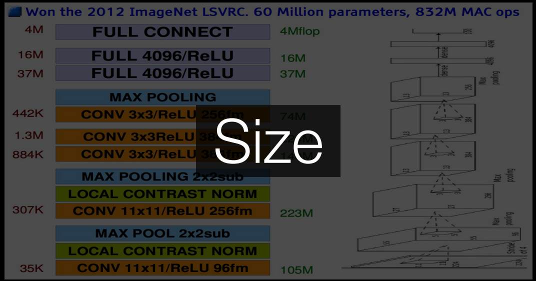


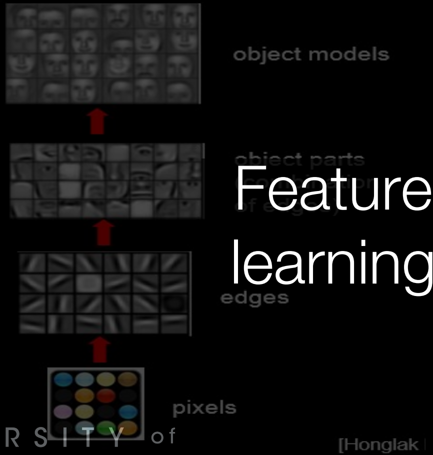
Dual CPU server, Intel E5-2698 v3 @ 2.3 GHz, 256 GB System Memory, Pre-Production Tesla P100

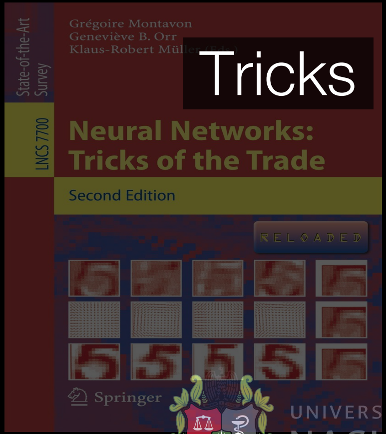
UNIVERSIDAD NACIONAL DE COLOMBIA

Deep learning recipe







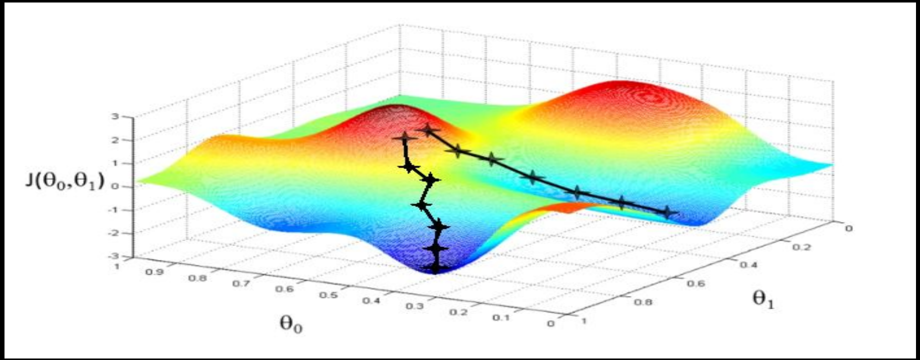


UNIVERSITY of HOUSTON

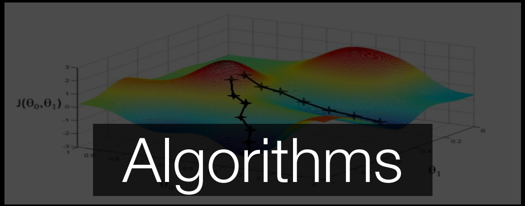
UNIVERSIDAD NACIONAL DE COLOMBIA

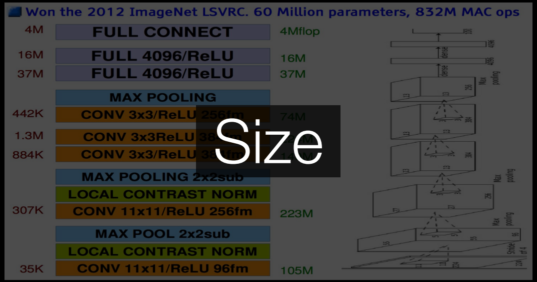
Algorithms

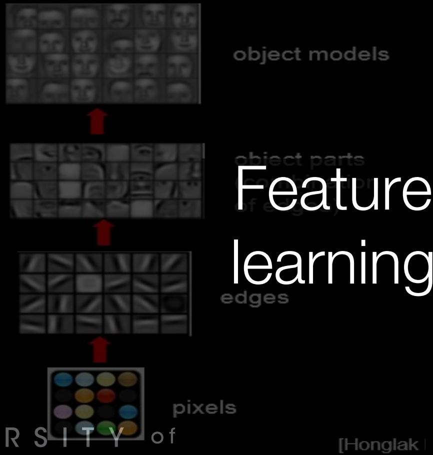
- Backpropagation
- Backpropagation through time
- Online learning (stochastic gradient descent)
- Softmax

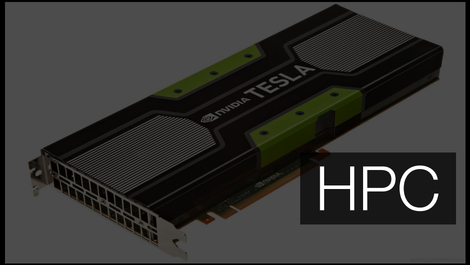


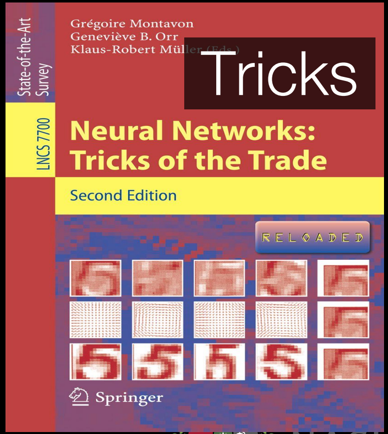
Deep learning recipe





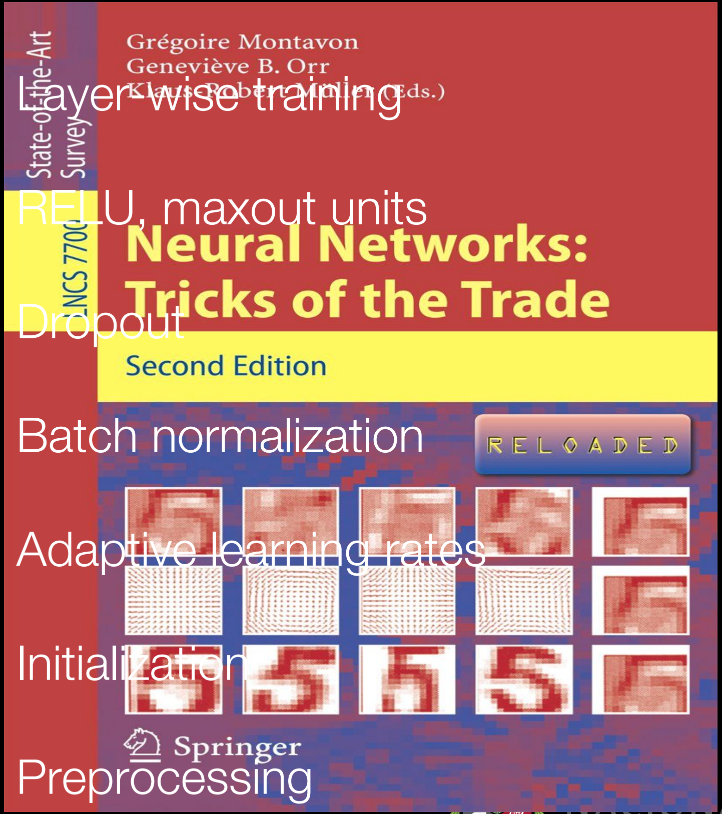






Tricks

- DL is mainly an engineering problem
- DL networks are hard to train
- Several tricks product of years of experience



- Layer-wise training
- RELU, maxout units
- Dropout
- Batch normalization
- Adaptive learning rates
- Initialization
- Preprocessing

UNIVERSITY of HOUSTON

DE COLOMBIA

Applications

- Computer vision:
    - Image: annotation, detection, segmentation, captioning
    - Video: object tracking, action recognition, segmentation
- Speech recognition and synthesis
- Text: language modeling, word/text representation, text classification, translation
- Biomedical image analysis