Deep learning

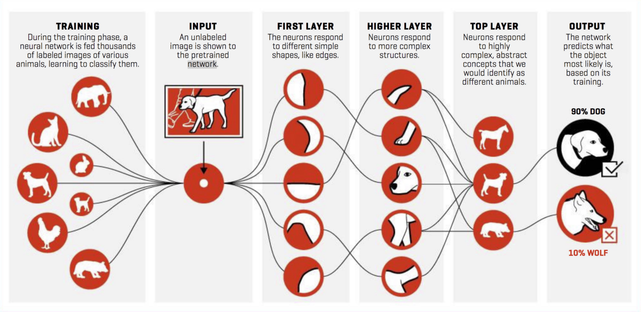
Stéphane Canu
asi.insa-rouen.fr/enseignants/~scanu

2016, Fréjus

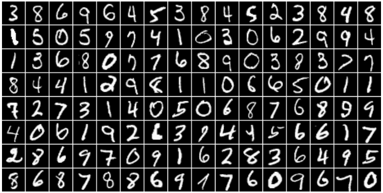
September 30, 2016

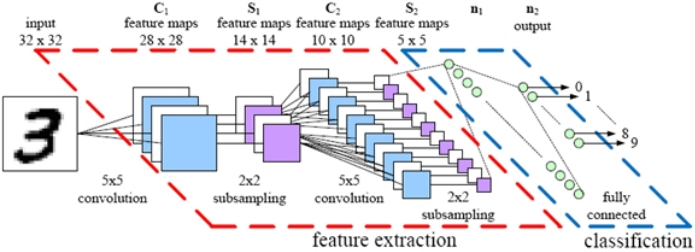
Plan

1. Apprendre à étiqueter une image avec un réseaux de neurones
   - The ImageNet challenge
   - Deep Learning architecture
   - Deep network training


TRAINING
During the training phase, a neural network is fed thousands of labeled images of various animals, learning to classify them.
INPUT
An unlabeled image is shown to the pretrained network.
FIRST LAYER
The neurons respond to different simple shapes, like edges.
HIGHER LAYER
Neurons respond to more complex structures.
TOP LAYER
Neurons respond to highly complex, abstract concepts that we would identify as different animals.
OUTPUT
The network predicts what the object most likely is, based on its training.
90% DOG
10% WOLF

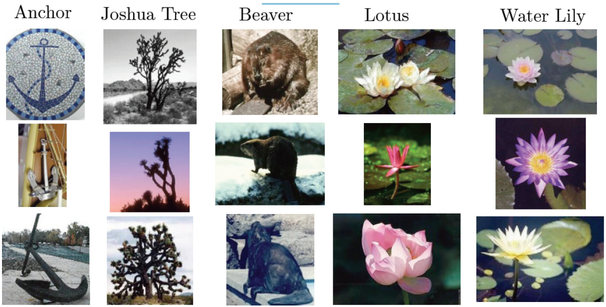
OCR: the MNIST database (Y. LeCun, 1989)



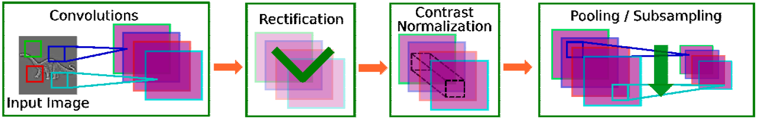


use convolution layers

The caltech 101 database (2004)



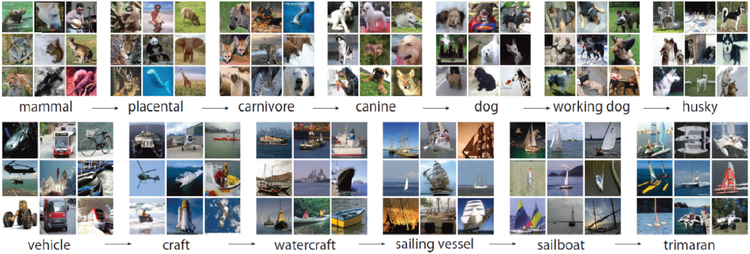
- 101 classes,
- 30 training images per category
- ...and the winner is NOT a deep network
  - dataset is too small



use convolution + Recitification + Normalization + Pooling

in What is the Best Multi-Stage Architecture for Object Recognition? Jarrett et al, 2009

The image net database (Deng et al., 2012)

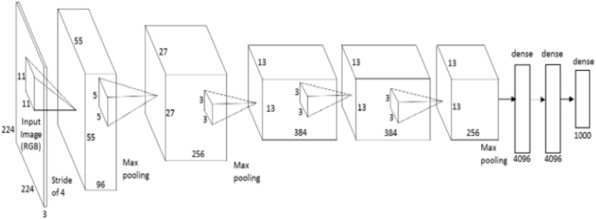

mammal → placental → carnivore → canine → dog → working dog → husky
vehicle → craft → watercraft → sailing vessel → sailboat → trimaran

ImageNet = 15 million labeled high-resolution images of 22,000 categories.
Large-Scale Visual Recognition Challenge (a subset of ImageNet)

- 1000 categories.
- 1.2 million training images,
- 50,000 validation images,
- 150,000 testing images.

www.image-net.org/challenges/LSVRC/2012/

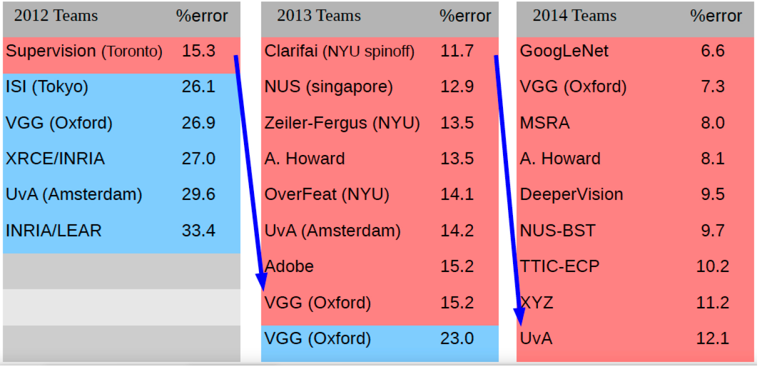
Deep architecture and the image net



The *Alex Net* architecture [Krizhevsky, Sutskever, Hinton, 2012]

- 60 million parameters
- using 2 GPU
- regularization
  - data augmentation
  - dropout
  - weight decay

A new fashion in image processing



2012 Teams	%error
Supervision (Toronto)	15.3
ISI (Tokyo)	26.1
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

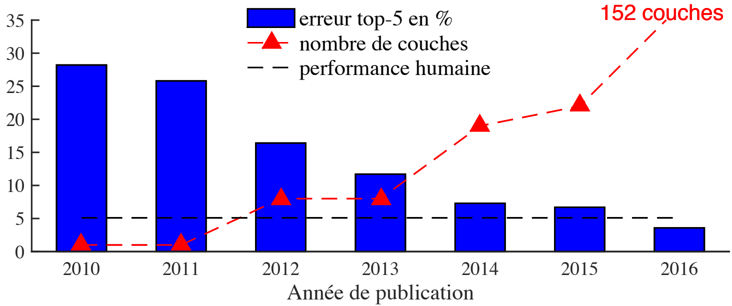
2013 Teams	%error
Clarifai (NYU spinoff)	11.7
NUS (singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	15.2
VGG (Oxford)	23.0

2014 Teams	%error
GoogLeNet	6.6
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	12.1

🟦 shallow approaches

🟥 deep learning

ImageNet results



2012 Alex Net

2013 ZFNet

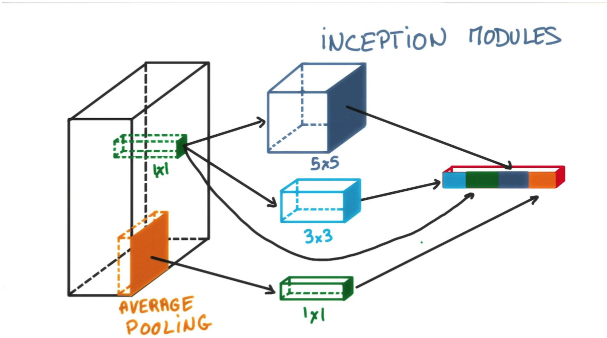
2014 VGG

2015 GoogLeNet / Inception

2016 Residual Network

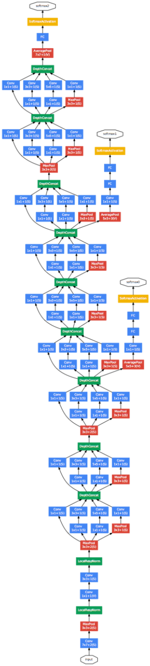
karpathy's blog: `karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/`

From 15% to 7%: Inceptionism

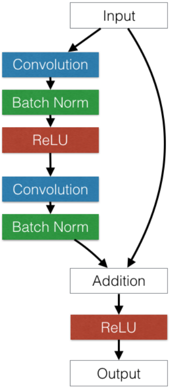


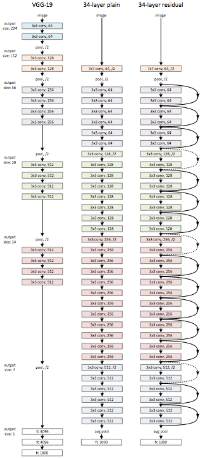
Network in a network (deep learning lecture at Udacity)

Christian Szegedy et. al. Going deeper with convolutions. CVPR 2015.



From 7% to 3%: Residual Nets

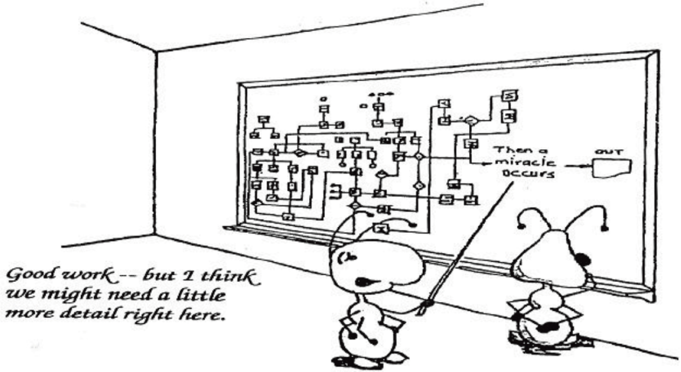




Beating the gradient vanishing effect

K. He *et al*, 2016

Learning Deep architecture


Then a miracle occurs
OUT
Good work -- but I think we might need a little more detail right here.

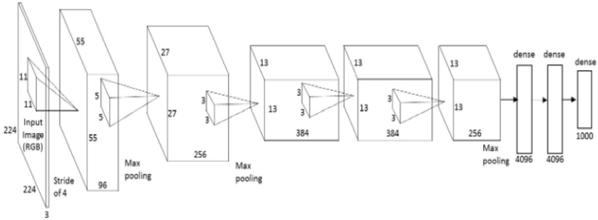
$$\min_{W \in \mathbb{R}^d} \sum_{i=1}^{n} \|f(x_i, W) - y_i\|^2 + \lambda \|w\|^2$$

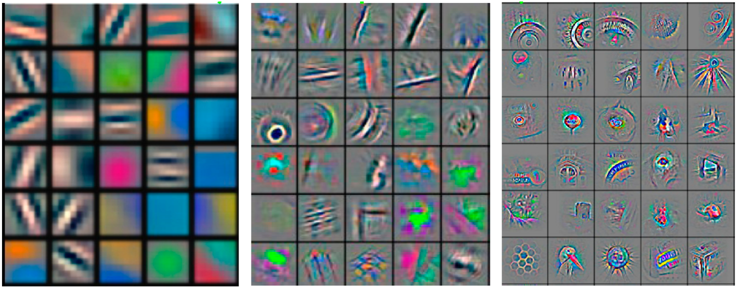
- $d = 60 \times 10^6$
- $n = 1,200\,000$ ++
- $\lambda = 0.0005$

f is a deep NN

Y. Bengio tutorial

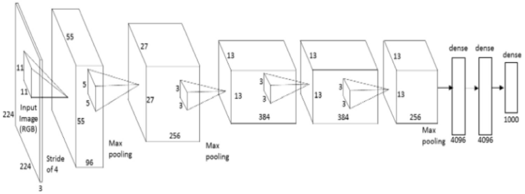
Then a m. occurs: learning internal representation

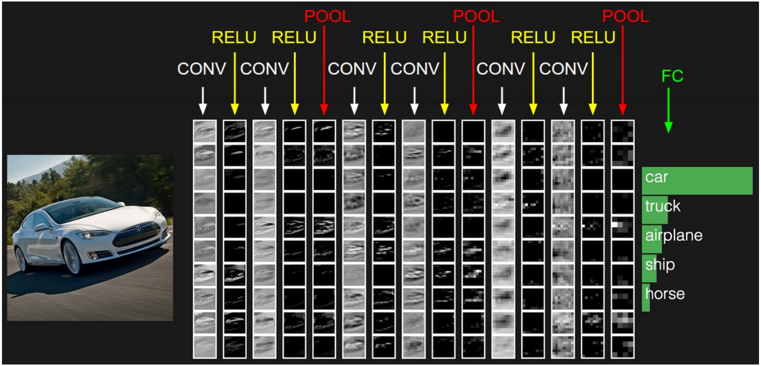




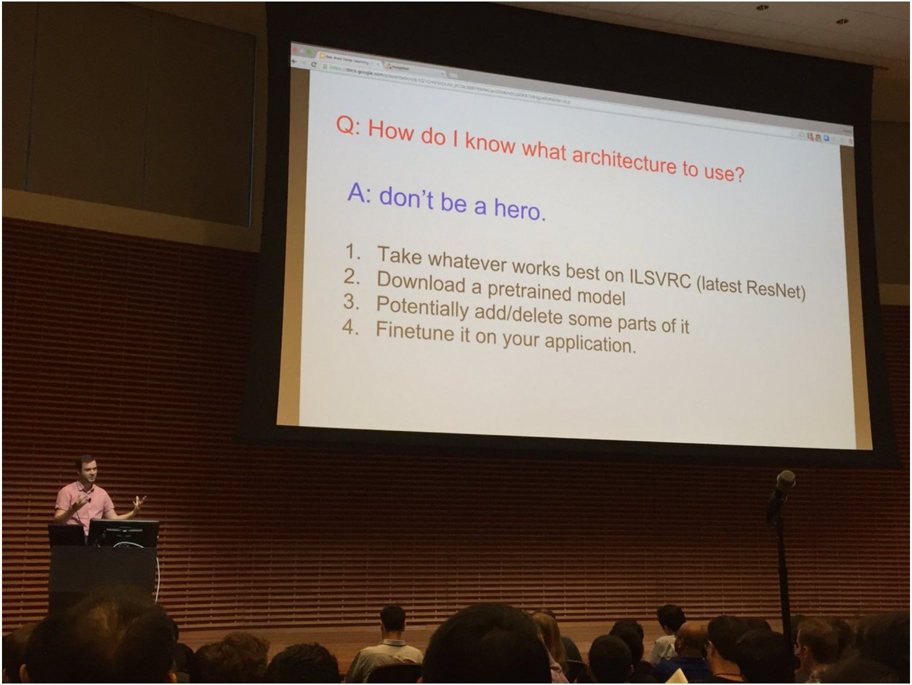
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Then a m. occurs: learning internal representation





CS231n: Convolutional Neural Networks for Visual Recognition

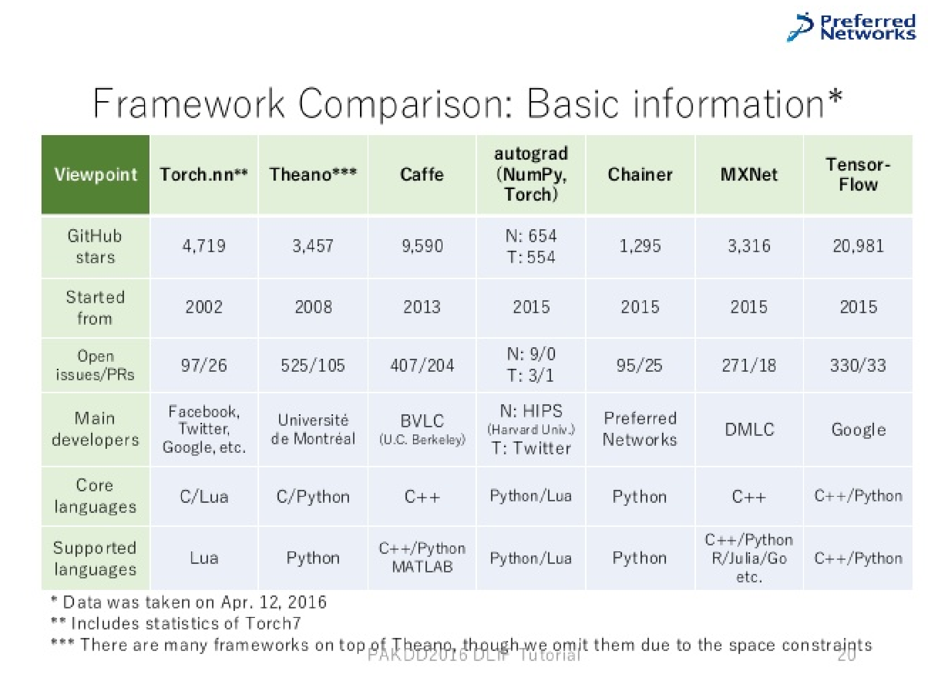


Q: How do I know what architecture to use?

A: don't be a hero.

1. Take whatever works best on ILSVRC (latest ResNet)
2. Download a pretrained model
3. Potentially add/delete some parts of it
4. Finetune it on your application.

Andrej Karpathy, Deep Learning Summer School 2016



Framework Comparison: Basic information*

Viewpoint	Torch.nn**	Theano***	Caffe	autograd (NumPy, Torch)	Chainer	MXNet	Tensor-Flow
GitHub stars	4,719	3,457	9,590	N: 654 T: 554	1,295	3,316	20,981
Started from	2002	2008	2013	2015	2015	2015	2015
Open issues/PRs	97/26	525/105	407/204	N: 9/0 T: 3/1	95/25	271/18	330/33
Main developers	Facebook, Twitter, Google, etc.	Université de Montréal	BVLC (U.C. Berkeley)	N: HIPS (Harvard Univ.) T: Twitter	Preferred Networks	DMLC	Google
Core languages	C/Lua	C/Python	C++	Python/Lua	Python	C++	C++/Python
Supported languages	Lua	Python	C++/Python MATLAB	Python/Lua	Python	C++/Python R/Julia/Go etc.	C++/Python

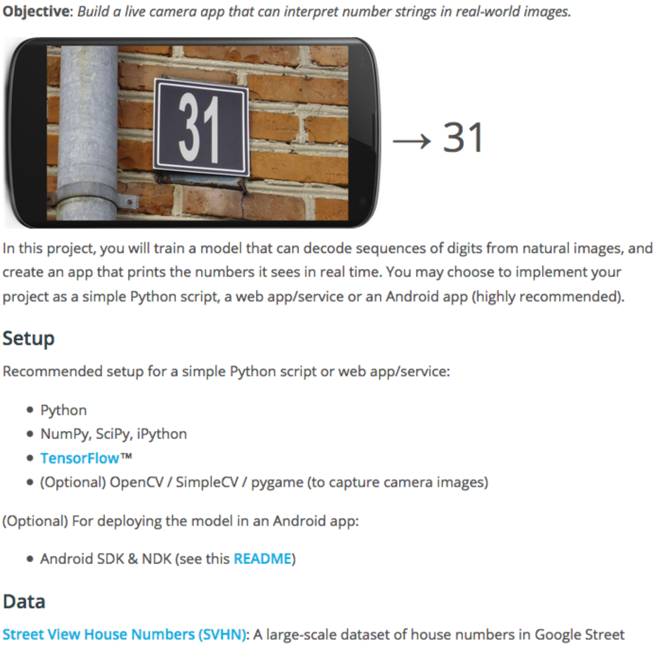
* Data was taken on Apr. 12, 2016

** Includes statistics of Torch7

*** There are many frameworks on top of Theano, though we omit them due to the space constraints

...and Keras (based on theano)

Pour en savoir plus : deep learning at Udacity (free course)



Objective: *Build a live camera app that can interpret number strings in real-world images.*

In this project, you will train a model that can decode sequences of digits from natural images, and create an app that prints the numbers it sees in real time. You may choose to implement your project as a simple Python script, a web app/service or an Android app (highly recommended).

Setup

Recommended setup for a simple Python script or web app/service:

- Python
- NumPy, SciPy, iPython
- TensorFlow™
- (Optional) OpenCV / SimpleCV / pygame (to capture camera images)

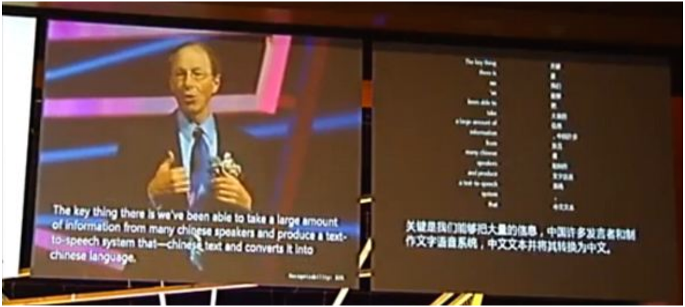
(Optional) For deploying the model in an Android app:

- Android SDK & NDK (see this README)

Data

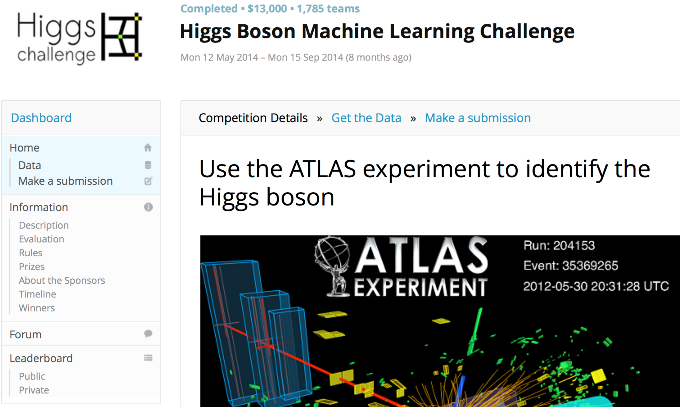
Street View House Numbers (SVHN): A large-scale dataset of house numbers in Google Street

Deep architecture: more contests and benchmark records


The key thing there is we've been able to take a large amount of information from many chinese speakers and produce a text-to-speech system that—chinese text and converts it into chinese language.
关键是我们能够把大量的信息，中国许多发言者和制作文字语音系统，中文文本并将其转换为中文。

- speech (phoneme) recognition and synthesis,
- text processing, automatic translation,
- image processing (MS COCO),
- ICDAR Chinese handwriting recognition benchmark,
- Grand Challenge on Mitosis Detection, Road sign recognition
- Higgs boson challenge

Machine Learning in High Energy Physics



Completed • $13,000 • 1,785 teams

Higgs Boson Machine Learning Challenge

Mon 12 May 2014 – Mon 15 Sep 2014 (8 months ago)

Competition Details » Get the Data » Make a submission

Use the ATLAS experiment to identify the Higgs boson

The winning model is "brute force"

- a bag of 70 dropout neural networks
- three hidden layers of 600 neurons
- produced by 2-fold stratified cross-validations

`https://www.kaggle.com/c/higgs-boson`

Comment ça marche ?

- Programmation par l'exemple (et beaucoup - des méga données)
- une bonne architecture :
  - Inception
  - Residual Networks
  - Ensembles
- un bon critère :
  - intégrer des connaissances a priori
  - adjustment aux données (énergie)
  - des mécanismes de régularisation
    - Drop out
    - Noise injection
- une bonne procédure d'optimisation
- des ressources informatiques suffisantes