Machine Learning, deep learning and optimization in Computer Vision

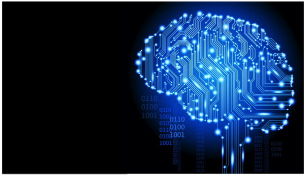
Stéphane Canu, Normandy University

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

QCAV 2017, Tokyo

May 15, 2017

Road map



My research time line

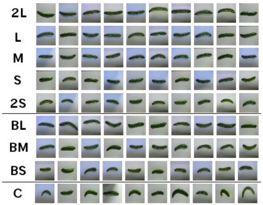
1985	Neural Networks (non linear PCA data analysis)	universal data fitting
1990	Machine Learning (understand regularization)	generalize
1995	SVM (convex optimization)	learning is optimization
2000	Model selection (automatic tuning)	tune hyperparameters
2005	MKL (kernel design)	automatized learning
2010	Dictionary learning (non convex optimization)	representation learning
2015	Deep learning (`www.deepinfrance.fr`)	without the magic

Success story: Koikes' cucumber farm

problem: sort cucumbers according to their quality

previously: sorted by hand

new solution: use data & machine learning (deep learning)



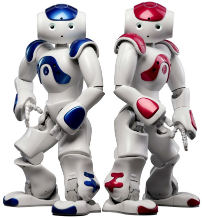
Other success stories in this morning talks

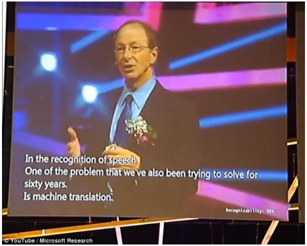
https://cloud.google.com/blog/big-data/2016/08/how-a-japanese-cucumber-farmer-is-using-deep-learning-and-tensorflow

Human vs. machine learning

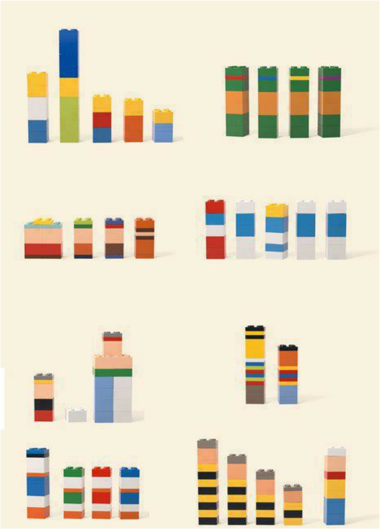
Child's learning capacities

- to walk: one year
- to speak: two years
- to think: the rest of my life





Learning to think



Machine learning definition

Machine Learning (T. Mitchell, 2006)

A computer program CP learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E

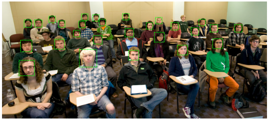
Key points

- experience E : data
- performance measure P : optimization
- tasks T : utility
  - automatic translation
  - play chess or go
  - ...do what humans do

Exemples of image processing related learning tasks T

Recognize

Detect



Localize

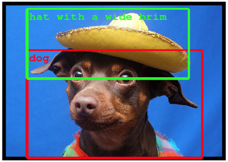
Predict

$\Rightarrow$

Read

Avenue des Sapins

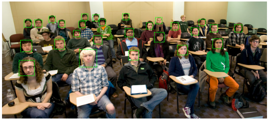
Scene analysis



Exemples of image processing related learning tasks T

Recognize

Detect



Localize

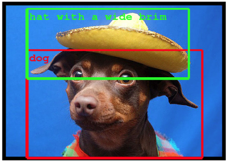
Predict

$\Rightarrow$

Read

Avenue des Sapins

Scene analysis



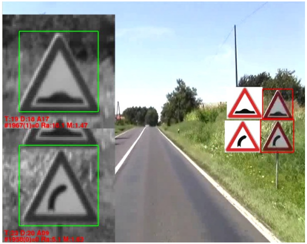
A single device can solve them all:

$\rightarrow$ our brain

Detection, tracking and recognition of traffic signs (2011-13)

Recognition German Traffic Sign Recognition Benchmark (GTSRB) data set, containing 51839 labelled images of real-world traffic signs.

Detection The German Traffic Sign Detection Benchmark is a single-image detection assessment 900 images (devided in 600 training images and 300 evaluation images)



and the winner is

→ Deep learning gives very good results on both tasks

The traditional machine learning flow

(Raw image)

preprocessing

(normalized image)

feature extraction

(features: HOG, SIFT...)

prediction

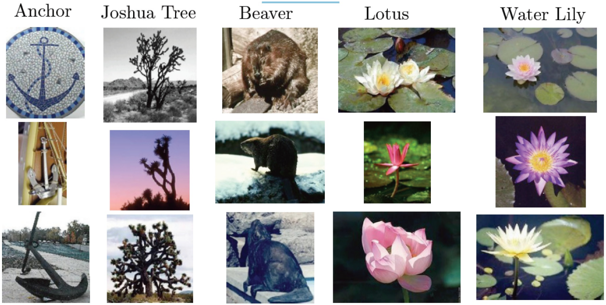
(decision)

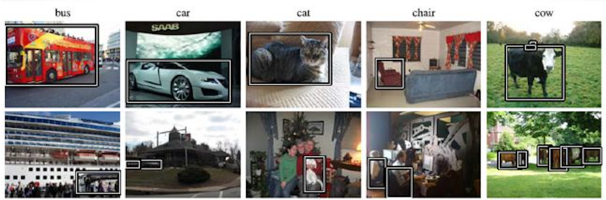
cf. Kaho Yamada & Yusuke Fujita's QCAV'17 talks

With good features...

...machine learning is easy

From caltech 101 database (2004) to VOC (2010)





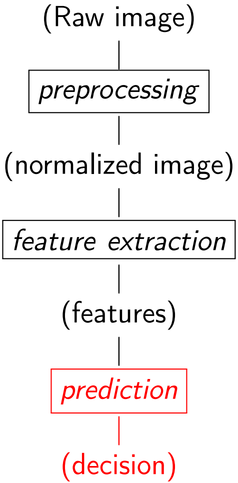
- 101 to 20 classes
- 30 to 1000 training images per category
- Perronnin et al. 2010,
  - hand-crafted Features
    - ★ SIFT
    - ★ Fisher Vectors
    - ★ Deformable Parts Pooling
  - SVM learning algo

Good features. . .

. . . provide good results

Hand-crafted features limitations

- they are application dependent
- they need a lot of effort
- they don't scale



Hand-crafted features limitations

- they are application dependent
- they need a lot of effort
- they don't scale

(Raw image)

preprocessing

(normalized image)

feature extraction

(features)

prediction

(decision)

Learn the features...

...together with the prediction

So far so good

- Machine learning is Example-based programming
  - task
  - data
  - performance
- A lot of applications in computer vision (as human learn to see)
  - detection
  - recognition
  - localization
  - captioning...
- Deep learning seems to be the future of machine learning
  - the dream of using a single multi task device for all vision tasks
  - on many application deep learning give the best results
  - feature learning is a challenge addressed by deep learning

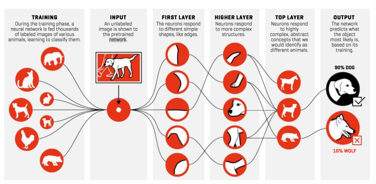
Road map

1. Machine learning in computer vision

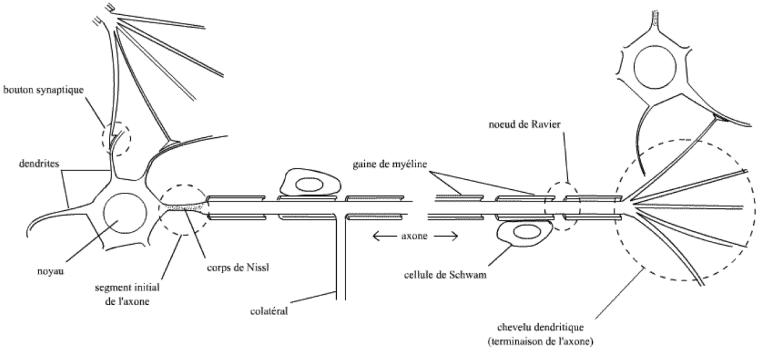
2. Deep learning
   - From neural networks to deep learning
   - ImageNet
   - The deep fashion

3. What's new in deep learning?
   - Big is beautiful
   - Learning and optimization
   - Deep architectures

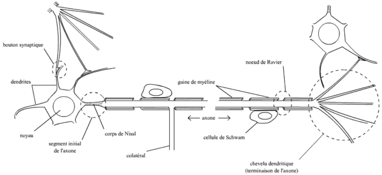
4. Conclusion

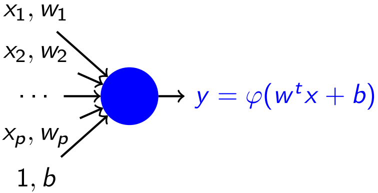


The biological neuron

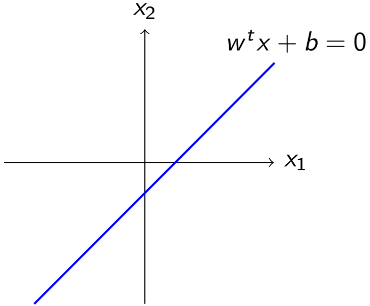


The formal neuron (McCulloch & Pitts, 1943)



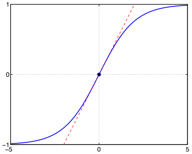


Defines a hyperplane



- x input $\in \mathbb{R}^p$
- w weight, b bias
- φ activation function
- y output $\in \mathbb{R}$

$\varphi(t) = \tanh(t)$



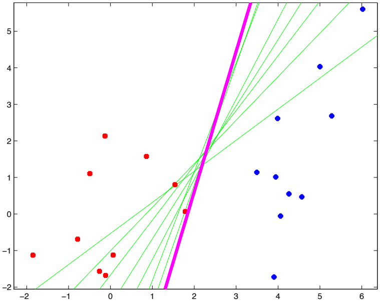
The formal neuron as a learning machine (Perceptron 1958)

Fitting the data for the 2 classes classification problem

given n pairs of input–output data $\mathbf{x}_i, y_i, i = 1, n$
find w such that $\underbrace{\varphi(\mathrm{w}^t\mathbf{x}_i)}_{\text{prediction of the model}} = \underbrace{y_i}_{\text{ground truth}}$

make some initial guess for w

- pick some data
- extract information
- improve w



what about new data – generalization?

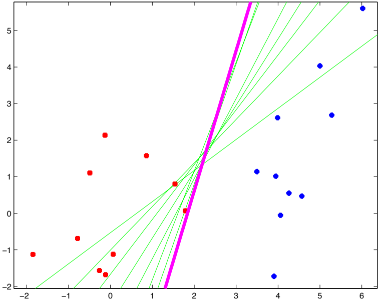
Fitting the data with an energy-based model

Fitting the data $$\min_{\mathrm{w}\in\mathbb{R}^{p+1}} \sum_{i=1}^{n} \big(\underbrace{\varphi(\mathrm{w}^t \mathbf{x}_i)}_{\text{prediction}} - \underbrace{y_i}_{\text{truth}} \big)^2$$

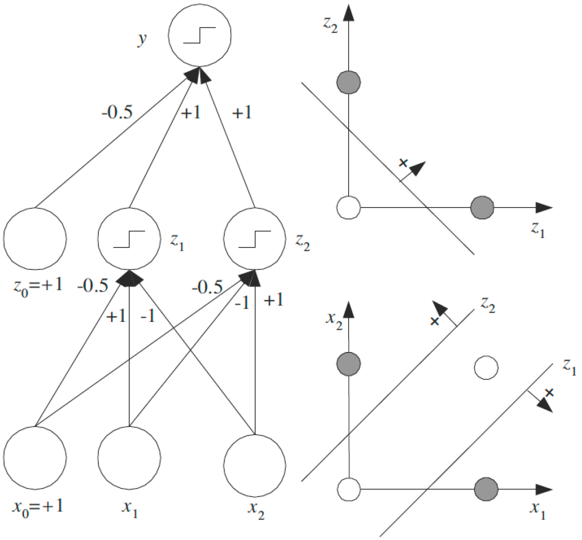
Improve w $$\mathrm{w}^{\text{new}} = \mathrm{w}^{\text{old}} - \rho \underbrace{(\varphi(w^t \mathbf{x}_i) - y_i)\, \nabla_{\mathrm{w}} \varphi(w^t \mathbf{x}_i)}_{\text{extracted information}}$$

Algorithm 1 Adaline (1960)

Data: w initialization, ρ fixed stepsize
Result: w
while *not converged* **do**
  $\mathbf{x}_i, y_i \leftarrow$ pick a point i at random
  $\mathbf{d} \leftarrow (w^t \mathbf{x}_i - y_i)\mathbf{x}_i$
  $\mathrm{w} \leftarrow \mathrm{w} - \rho \mathbf{d}$
  check that the cost decease
end



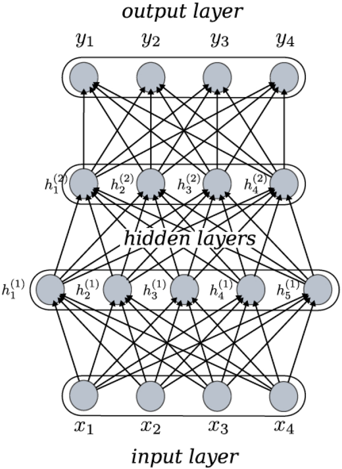
Non linearity combining linear neurons: the Xor case



Alpaydın, Introduction to Machine Learning, 2010

Neural networks as universal approximator

Running several neurons at the same time



$$y = \varphi(W_3 h^{(2)})$$

$$h^{(2)} = \varphi(W_2 h^{(1)})$$

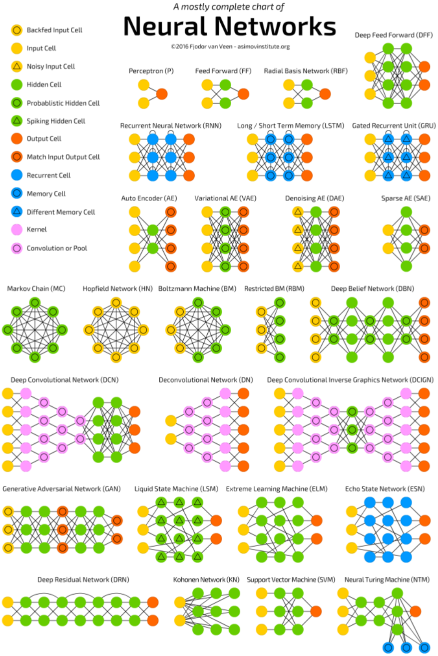
$$h^{(1)} = \varphi(W_1 \mathbf{x})$$

$$\mathbf{x}$$

Multilayered neural networks in layers

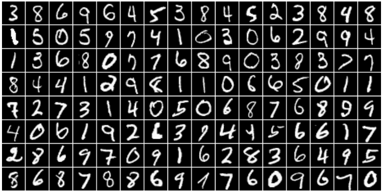
Use backpropagation to learn internal representation W_1, W_2, W_3

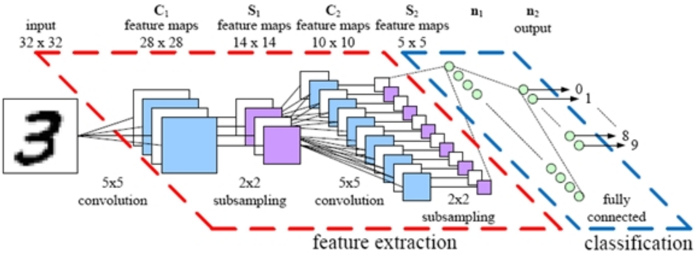
from L. Arnold PhD



The Asimov Institute: `http://www.asimovinstitute.org/neural-network-zoo/`

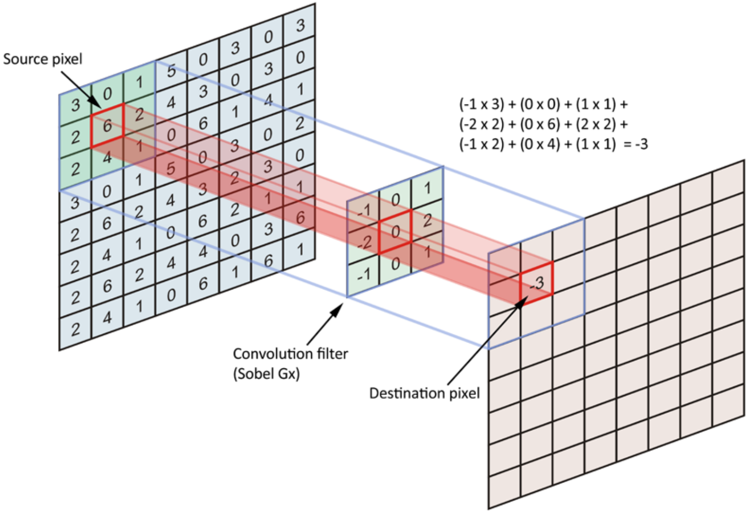
OCR: the MNIST database (Y. LeCun, 1989)





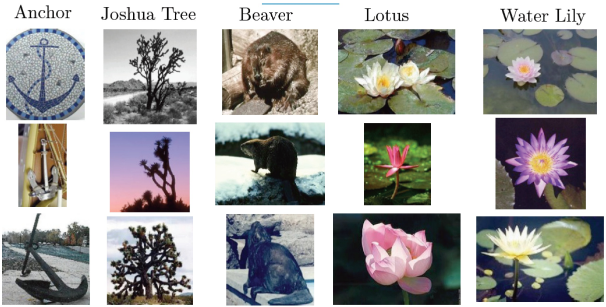
use convolution layers

What are convolutional neural networks (CNN)?

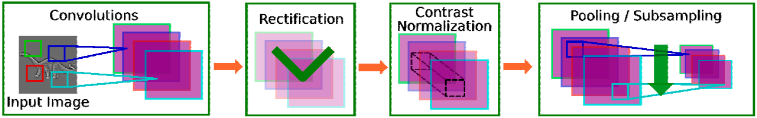


convolution reduces the number of parameters

The caltech 101 database (2004)



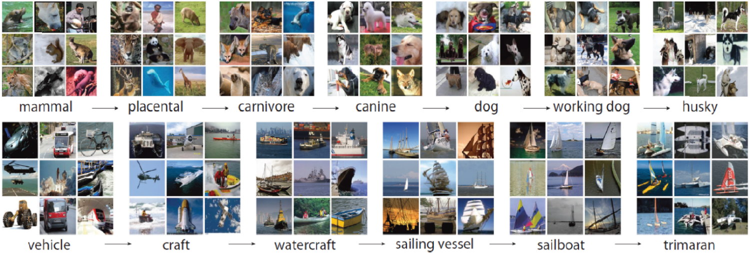
- 101 classes,
- 30 training images per category
- ...and the winner is NOT a deep network
  - dataset is too small



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)



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/

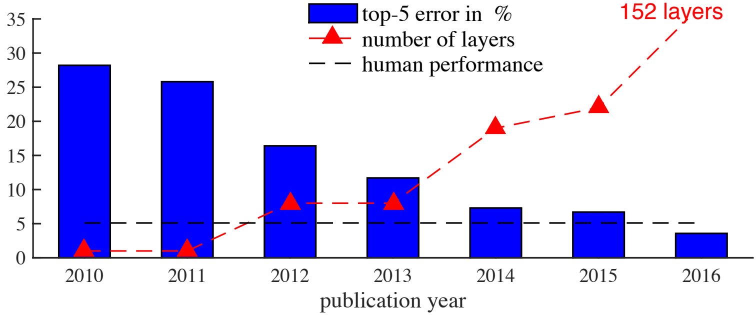
A new fashion in image processing

2012 Teams	%error	2013 Teams	%error	2014 Teams	%error
Supervision (Toronto)	15.3	Clarifai (NYU spinoff)	11.7	GoogLeNet	6.6
ISI (Tokyo)	26.1	NUS (singapore)	12.9	VGG (Oxford)	7.3
VGG (Oxford)	26.9	Zeiler-Fergus (NYU)	13.5	MSRA	8.0
XRCE/INRIA	27.0	A. Howard	13.5	A. Howard	8.1
UvA (Amsterdam)	29.6	OverFeat (NYU)	14.1	DeeperVision	9.5
INRIA/LEAR	33.4	UvA (Amsterdam)	14.2	NUS-BST	9.7
		Adobe	15.2	TTIC-ECP	10.2
		VGG (Oxford)	15.2	XYZ	11.2
		VGG (Oxford)	23.0	UvA	12.1

shallow approaches

deep learning

ImageNet results



2012 Alex Net

2013 ZFNet

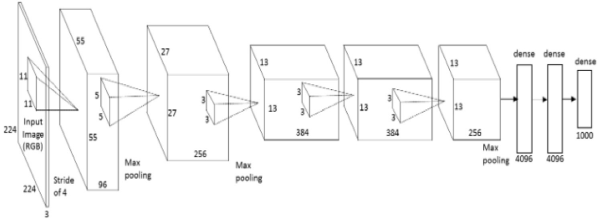
2014 VGG

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/`

Deep architecture and the image net (15%)

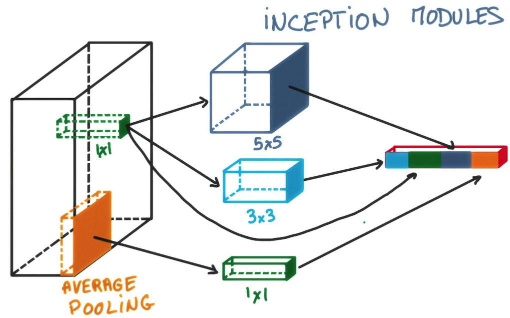


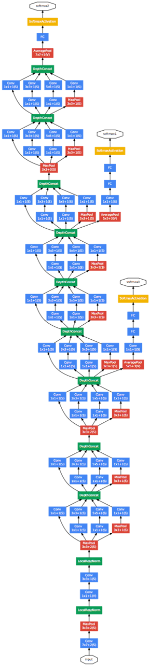
The *Alex Net* architecture [Krizhevsky, Sutskever, Hinton, 2012]

Convolution + Recitification (ReLU) + Normalization + Pooling

- 60 million parameters
- using 2 GPU – 6 days
- regularization
  - data augmentation
  - dropout
  - weight decay

From 15% to 7%: Inceptionism

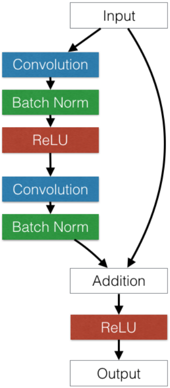


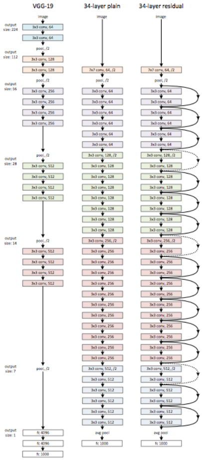


Network in a network (deep learning lecture at Udacity)

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

From 7% to 3%: Residual Nets

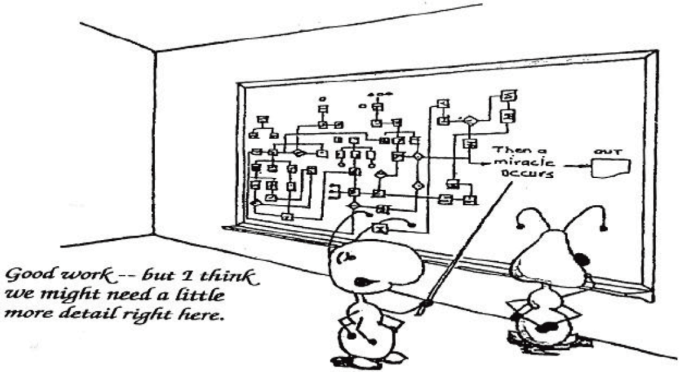




Beating the gradient vanishing effect

K. He *et al*, 2016

Learning Deep architecture

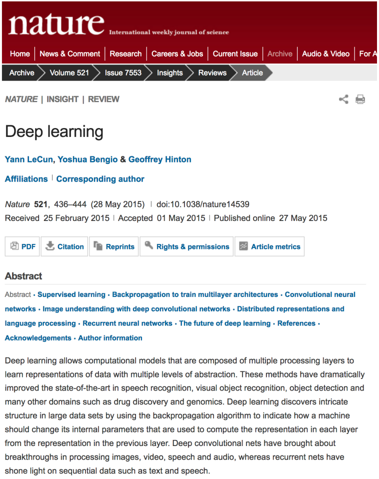


$$\min_{W \in \mathbb{R}^d} \sum_{i=1}^{n} \|f(x_i, W) - y_i\|^2 + \lambda \|w\|^2$$

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

f is a deep NN

The deep learning research bubble



nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For A

Archive > Volume 521 > Issue 7553 > Insights > Reviews > Article

NATURE | INSIGHT | REVIEW

Deep learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton

Affiliations | Corresponding author

Nature **521**, 436–444 (28 May 2015) | doi:10.1038/nature14539

Received 25 February 2015 | Accepted 01 May 2015 | Published online 27 May 2015

PDF | Citation | Reprints | Rights & permissions | Article metrics

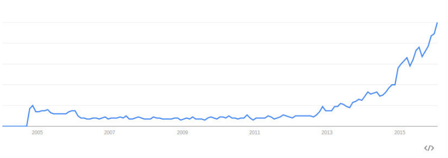
Abstract

Abstract · Supervised learning · Backpropagation to train multilayer architectures · Convolutional neural networks · Image understanding with deep convolutional networks · Distributed representations and language processing · Recurrent neural networks · The future of deep learning · References · Acknowledgements · Author information

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech.

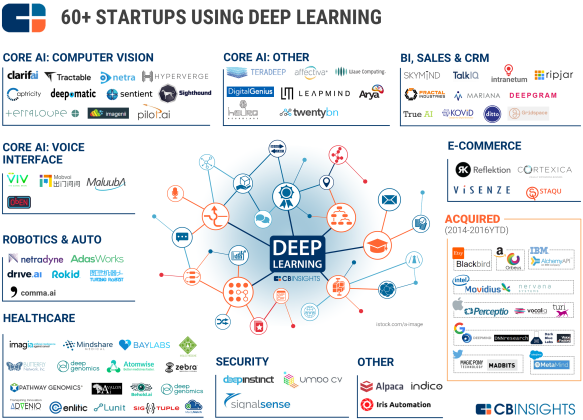
- Paper by "Deep Learning Conspiracy" in Nature, 2015 (2500 citations)
- CVPR is becoming deep

  . . . try a ConvNet-based baseline
- NIPS attendance from 500 to 5000
- Google Trends for the search term "deep learning"



Deap learning applications

- image
- audio
- speech
- text (NLP)
- translation
- robotics
- playing games
- science (Higgs Boson)



Mostly related with

low level perception

Deep learning and the industry

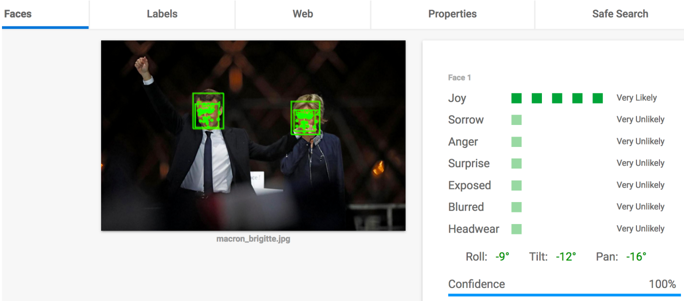
- deep learning startup
- data science, Artificial intelligence and deep learning
- the GAFA
  - they got the infrastructure (hard+software)
  - they got the data
  - deep learning bridges the gap between applications and ML

Deep Learning As A Service

a lot of available API (google, microsoft, Nvidia, Amazon...)

Google cloud plateform https://cloud.google.com/vision/

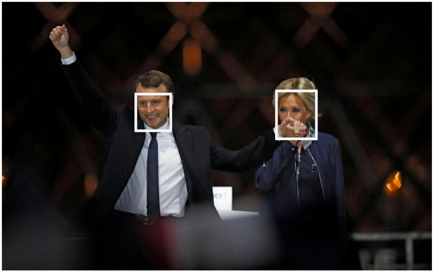
Google cloud plateform https://cloud.google.com/vision/



Google cloud deep learning engine

democratizing access to the world's most powerful deep learning systems

Watson sees...



JSON

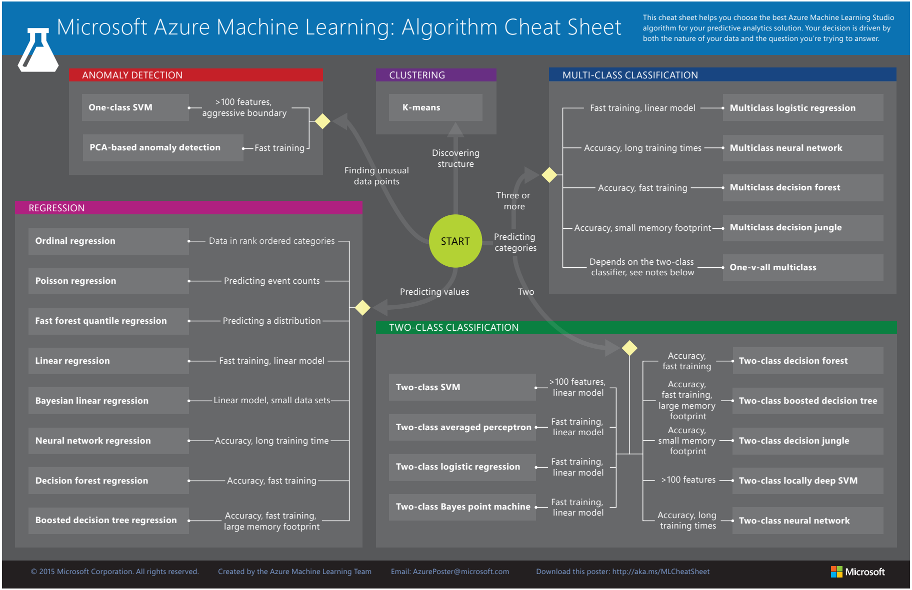
Classes	Score
politician	0.59
person	0.72
President of the United States	0.56
president	0.56
speaker	0.53
coal black color	0.99

JSON

Faces	Score
age 55 - 64	0.47
female	0.99
age 35 - 44	0.40
male	1.00

Did We Wow You? ◯ Yes ◯ No

the Microsoft Azure API, machine learning studio



The artificial intelligence (AI) renaissance (The Economist)

- form deep learning to artificial intelligence getting machines to solve problems now reserved for humans
- AI Partnership
  Amazon, Facebook, Google, IBM, and Microsoft
- AI initiative, AI white papers,
  Canada, US, Japan, France...
- AI institutes

FORTUNE SUBSCRIBE

By Roger Parloff

Illustration by Justin Metz

SEPTEMBER 28, 2016, 5:00 PM EDT

WHY DEEP LEARNING IS SUDDENLY CHANGING YOUR LIFE

Decades-old discoveries are now electrifying the computing industry and will soon transform corporate America.

Over the past four years, readers have doubtlessly noticed quantum leaps in the quality of a wide range of everyday technologies.

Most obviously, the speech-recognition functions on our smartphones work much better than they used to. When we use a voice command to call our spouses, we reach them now. We aren't connected to Amtrak or an angry ex.

In fact, we are increasingly interacting with our computers by just talking to them, whether it's Amazon's Alexa, Apple's Siri, Microsoft's Cortana, or the many voice-responsive features of Google. Chinese search giant Baidu says customers have tripled their use of its speech interfaces in the past 18 months.

Université de Montréal | Montreal Institute for Learning Algorithms

The current boom in AI...

...is really a boom in "deep learning" (The Economist)

So far so good

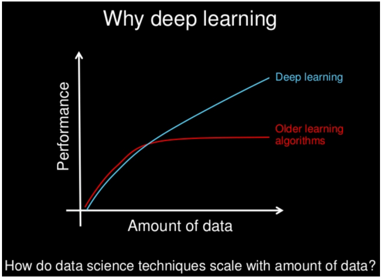
- from the formal neuron to deep learning
  - one neuron is a linear perceptron
  - many layered neurons are non linear multilayered perceptrons
  - deep networks is a new name for multilayered perceptrons
- deep learning breakthrough starts with ImageNet
  - better than human performances
  - on many perception tasks
- deep learning could transform almost any industry
  - the AI revolution

Neural networks+backpropagation exist since 1985

$\rightarrow$ what's new?

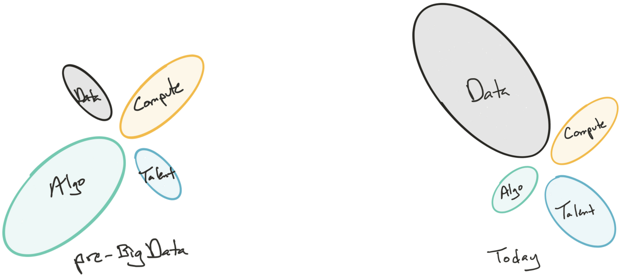
Road map

1. Machine learning in computer vision
2. Deep learning
   - From neural networks to deep learning
   - ImageNet
   - The deep fashion
3. What's new in deep learning?
   - Big is beautiful
   - Learning and optimization
   - Deep architectures
4. Conclusion



What's new with deep learning

- a lot of data (big data)
- big model (deep vs. shalow)
- big computing resources (hardware & software),
- and the talent to put it all together.



https://www.oreilly.com/ideas/the-four-dynamic-forces-shaping-ai

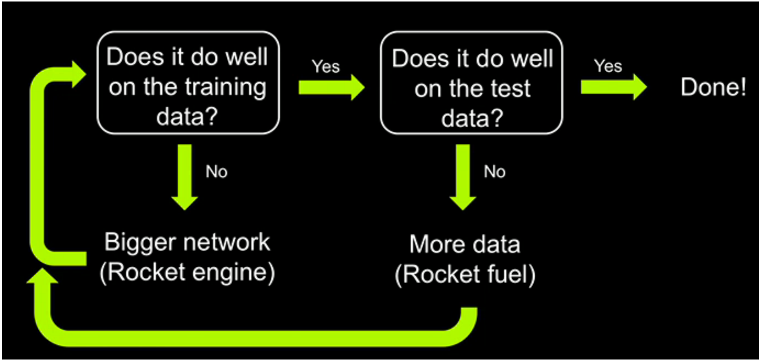
Dealing with a lot of data

- ImageNet: 1,200,000x256x256x3 (about 200GB) block of pixels
- MS COCO for supervised learning
  - Multiple objects per image
  - More than 300,000 images
  - More than 2 Million instances
  - 80 object categories
  - 5 captions per image
- YFCC100M for unsupervised learning
- Google Open Images, 9 million URLs to images annotated over 6000 categories
- Place recognition datasets (including Tokyo)

Andrew Ng basic recipe for machine learning

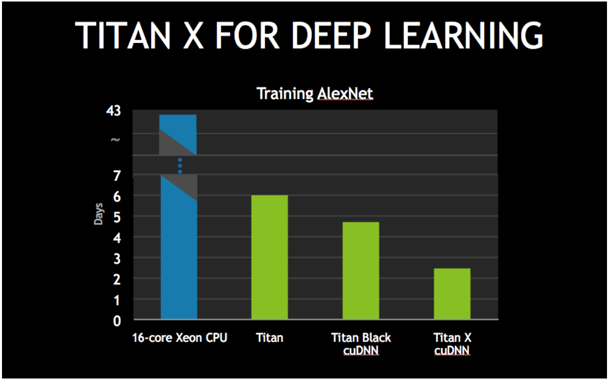
Why is Deep Learning taking off?

fuel = data engine = model (a deep network) .



Andrew Ng GTC 2015 Keynote, GPU Technology Nvidia

GPU needed



Now 2 hours with Nvidia DGX-1, and enough Memory

Yann LeCun:

learning a relevant model takes a 3 weeks

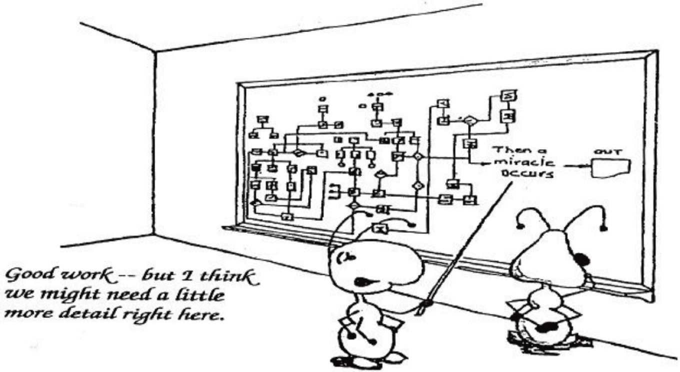
Deep learing frameworks

	Languages	Tutorials and training materials	CNN modeling capability	RNN modeling capability	Architecture: easy-to-use and modular front end	Speed	Multiple GPU support	Keras compatible
Theano	Python, C++	++	++	++	+	++	+	+
Tensor-Flow	Python	+++	+++	++	+++	++	++	+
Torch	Lua, Python (new)	+	+++	++	++	+++	++	
Caffe	C++	+	++		+	+	+	
MXNet	R, Python, Julia, Scala	++	++	+	++	++	+++	
Neon	Python	+	++	+	+	++	+	
CNTK	C++	+	+	+++	+	++	+	

Tensoflow (Google) is the most popular, Torch (facebook)

http://www.kdnuggets.com/2017/03/getting-started-deep-learning.html

Learning Deep architecture



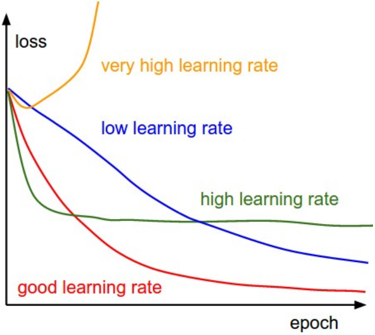
$$\min_{W \in \mathbb{R}^d} \sum_{i=1}^{n} \|f(x_i, W) - y_i\|^2 + \lambda \|w\|^2$$

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

f is a deep NN

New learning strategies

- adaptive learning rate
- regularization
- stochastic gradient acceleration (Bottou)
- noise injection (Bengio)
- drop out (Hinton)
- deep networks committees (Schmidhuber)
- auto encoder pre training (Bengio)



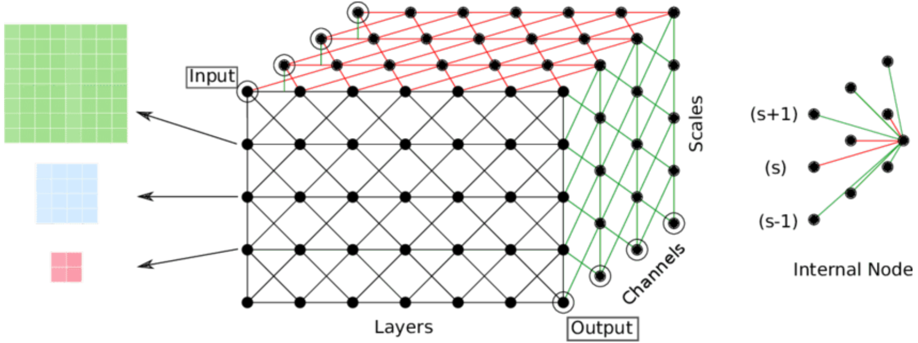
and many other

a very active research and engineering field

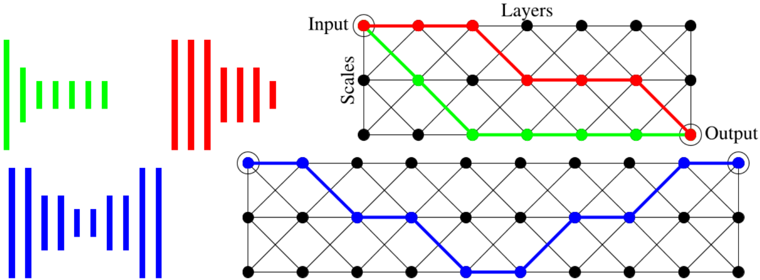
Learning deep architectures (1/2)

Convolutional Neural Fabrics (Saxena and Verbeek, NIPS 16)

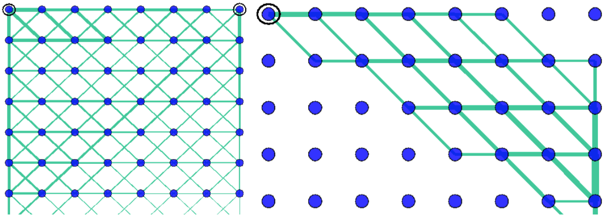
- problem: how to find the most relevant architecture
- todays solution: try and test
- A new solution: learn the architecture



Examples of deep architectures using neural fabrics (2/2)

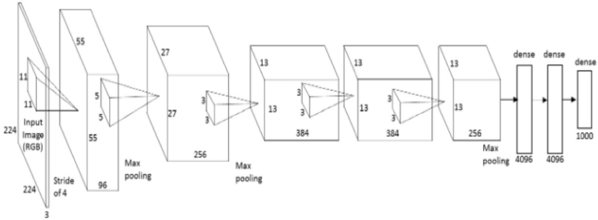


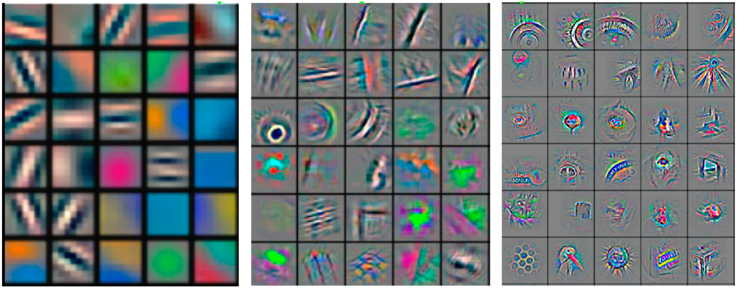
–Convolutional classifier –Convolutional classifier –autoencoder



Convolutional Neural Fabrics (Saxena and Verbeek, NIPS 16)

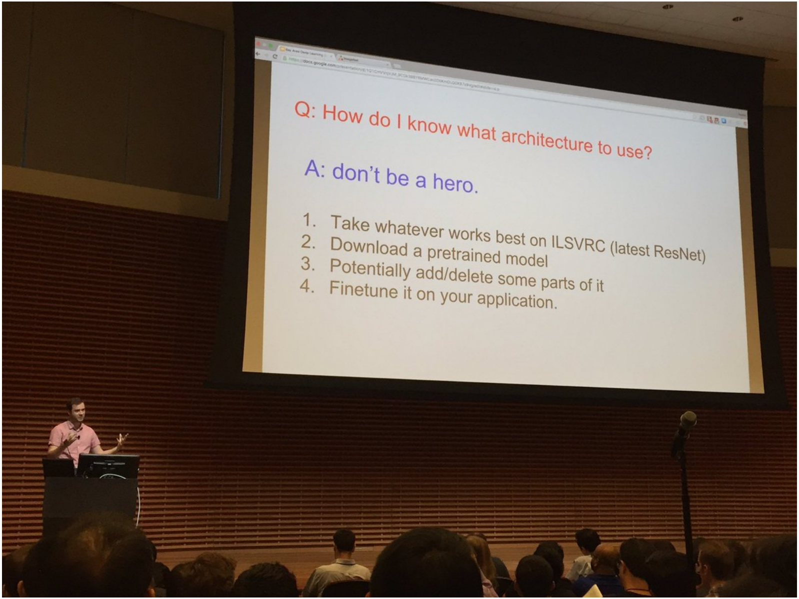
AlexNet works because of learning internal representation





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

How to start with deep learning?

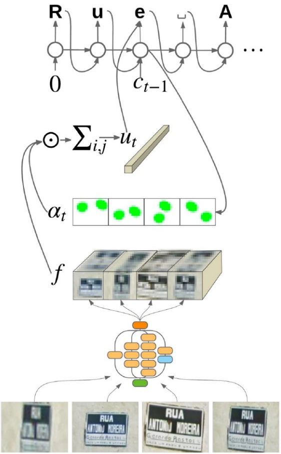


cf. Nathan Mundhenk's this morning talk
Andrej Karpathy, Deep Learning Summer School 2016

Success story: Updating Google Maps with Deep Learning

Requirements

- Installed TensorFlow library
- 158Gb to download FSNS dataset:
- 16Gb of RAM (32Gb is recommended)
- training $\sim$ 60 h with GPU Titan X
- to train from scratch: python train.py
- to train a model using a pre-trained inception weights as initialization:




```
wget http://download.tensorflow.org/models/inception_v3_2016_08_28.t
tar xf inception_v3_2016_08_28.tar.gz
python train.py --checkpoint_inception=inception_v3.ckpt
```

For more information: deep learning at Udacity (free course)

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

→ 31

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:

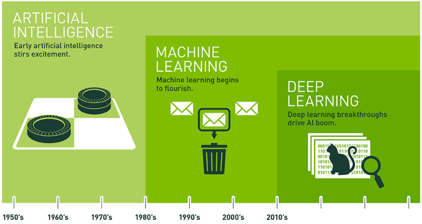
- Android SDK & NDK (see this README)

Data

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

Road map

1. Machine learning in computer vision
2. Deep learning
    - From neural networks to deep learning
    - ImageNet
    - The deep fashion
3. What's new in deep learning?
    - Big is beautiful
    - Learning and optimization
    - Deep architectures
4. Conclusion



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Conclusions

- summarize
  - deep learning major breakthrough with imagenet
  - GPUs, memory, existing framework, and a lot of data
  - if possible reuse deep networks (alexNet, VGG. . .)
- the future of deep learning
  - hard to compete with GAFAs'
  - theory needed
  - unsupervised learning
  - efficient learning regarding examples, time and energy
  - also I missed the GAN (Generative Adversarial Networks)
- and the future of machine learning?
  - learning to learn
  - green learning
  - optimization and learning

To go further

- books
  - I. Goodfellow, Y. Bengio & A. Courville, *Deep Learning*, MIT Press book, 2016 http://www.deeplearningbook.org/
  - Gitbook leonardoaraujosantos.gitbooks.io/artificial-inteligence/
- conferences
  - NIPS, ICLR, xCML, AIStats,
- Journals
  - JMLR, Machine Learning, Foundations and Trends in Machine Learning, machine learning survey http://www.mlsurveys.com/
- lectures
  - Deep Learning: Course by Yann LeCun at Collège de France in 2016 college-de-france.fr/site/en-yann-lecun/inaugural-lecture-2016-02-04-18h00.htm
  - Convolutional Neural Networks for Visual Recognition (Stanford)
  - deep mind (https://deepmind.com/blog/)
  - CS 229: Machine Learning at stanford Andrew Ng
- Blogs
  - Andrej Karpathy blog (http://karpathy.github.io/)
  - http://deeplearning.net/blog/
  - https://computervisionblog.wordpress.com/category/computer-vision/