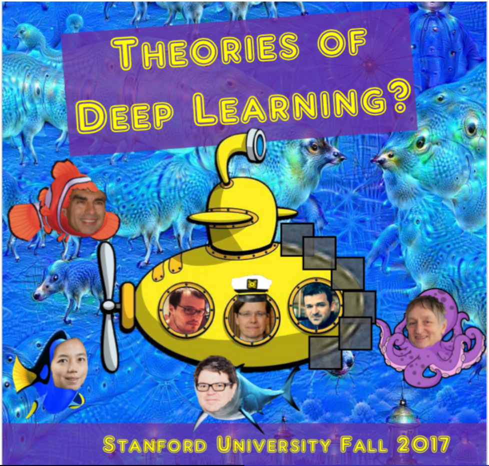
Lecture 01: Deep Learning Challenge: Is There Theory?

D Donoho/ H Monajemi/ V Papyan
Stats 385 Stanford

20170927

Stats 385 Fall 2017

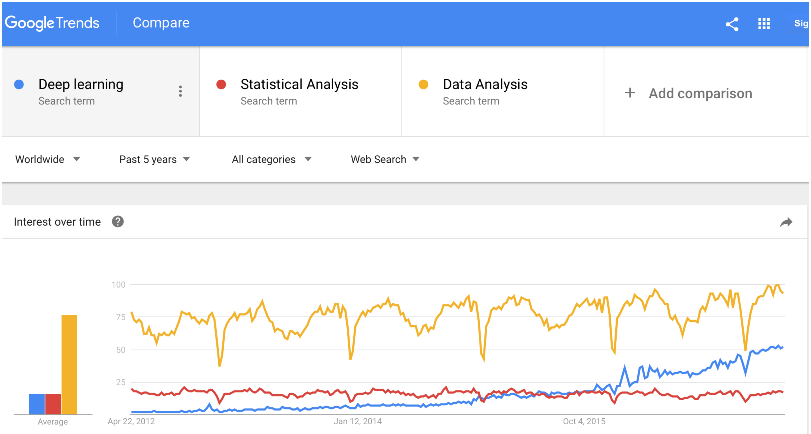


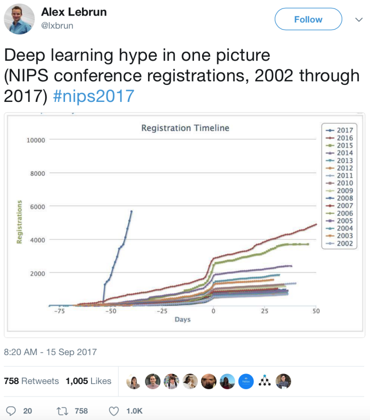
Outline

Disclaimer

This document contains images obtained by routine Google Images searches. Some of these images may perhaps be copyright. They are included here for educational noncommercial purposes and are considered to be covered by the doctrine of **Fair Use**. *In any event they are easily available from Google Images.*

It's not feasible to give full scholarly credit to the creators of these images. We hope they can be satisfied with the positive role they are playing in the educational process.


Google Trends
Compare
Sig
Deep learning
Search term
Statistical Analysis
Search term
Data Analysis
Search term
Add comparison
Worldwide
Past 5 years
All categories
Web Search
Interest over time
100
75
50
25
Average
Apr 22, 2012
Jan 12, 2014
Oct 4, 2015

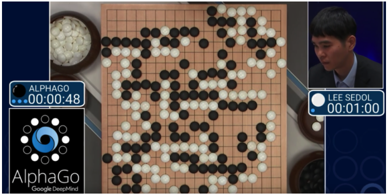

Alex Lebrun
@lxbrun
Follow
Deep learning hype in one picture
(NIPS conference registrations, 2002 through 2017) #nips2017
Registration Timeline
10000
8000
6000
4000
2000
Registrations
-75
-50
-25
0
25
50
Days
2017
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004
2003
2002
8:20 AM - 15 Sep 2017
758 Retweets 1,005 Likes
20
758
1.0K


Andrej Karpathy
@karpathy
Follow
Came to visit first class of @cs231n at Stanford. 2015: 150 students, 2016: 350, this year: 750. #aiinterestsingularity
12:11 PM - 4 Apr 2017
155 Retweets 623 Likes
19
155
623
michael_nielsen @michael_nielsen · Apr 4
Replying to @karpathy @cs231n
Faster than Moore's Law. At this rate - doubling each year - in 24 years everyone on Earth will be enrolled :-)

Reaching Human Level Performance

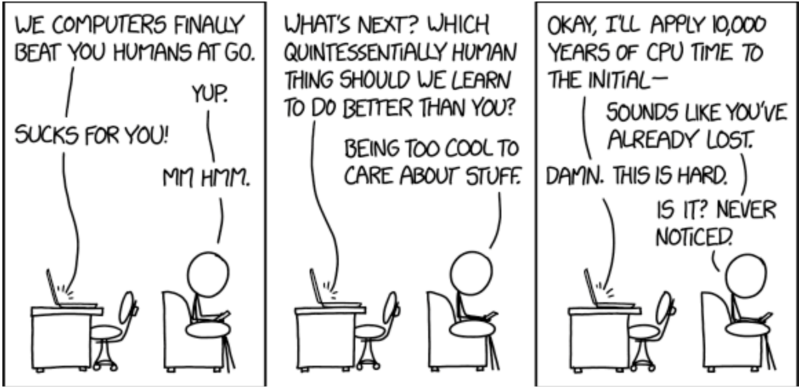
1997

2004



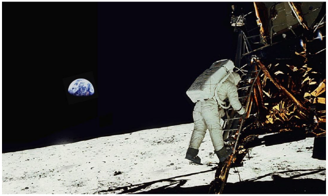
2017

Something DL Can't Do (per XKCD)



Landmark Applied Science – 20th century

Nuclear Power

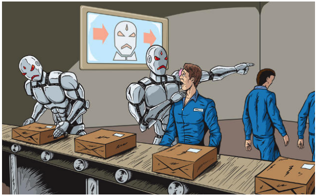


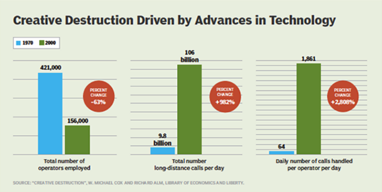
Powered Flight

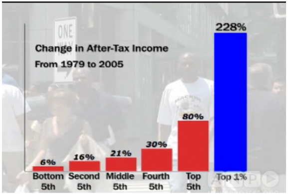
Antibiotics

Green Revolution

Impact on Humanity







Why this is happening now?

Key ingredients of DL have been in place for 25-30 years:

Landmark	Emblem	Epoch
Neocognitron	Fukushima	1980
CNN	Le Cun	mid 1980s'
Backprop	Hinton	mid 1980's
SGD	Le Cun, Bengio etc	mid 1990's
Various	Schmidhuber	mid 1980's

Some argue that no really new *ideas* were needed

Synchronies

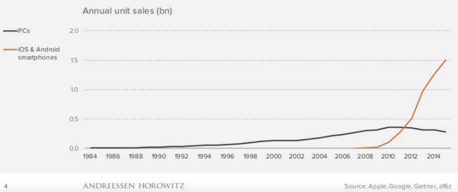
Over same timeframe – 2010-2014

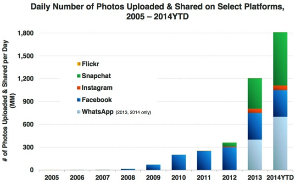
- Instagram, Snapchat emerge to global prominence
- Deep Learning catapults to global attention

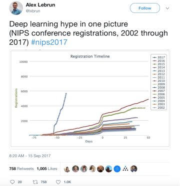
Coincides with emergence of

- Smartphone photography
- Cloud storage of selfie/smartphone photography

Happenings 2010-2014

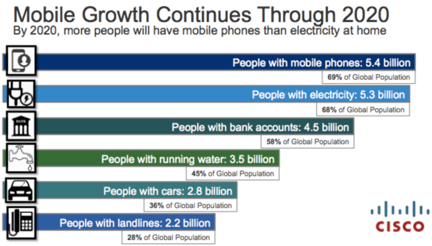


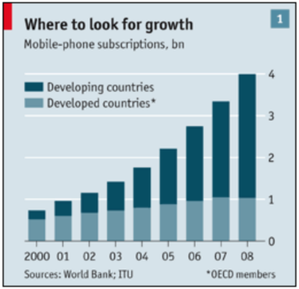




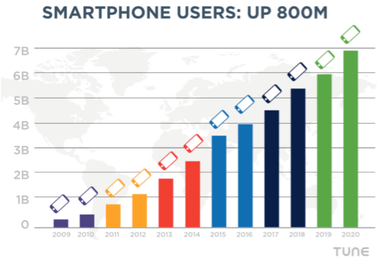
The Mobile Revolution

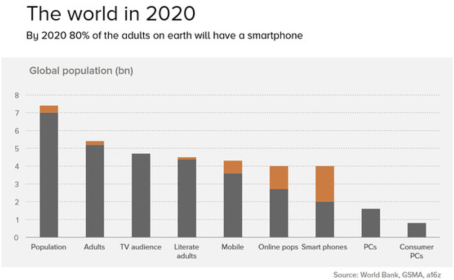
Mobile Is Spreading Everywhere



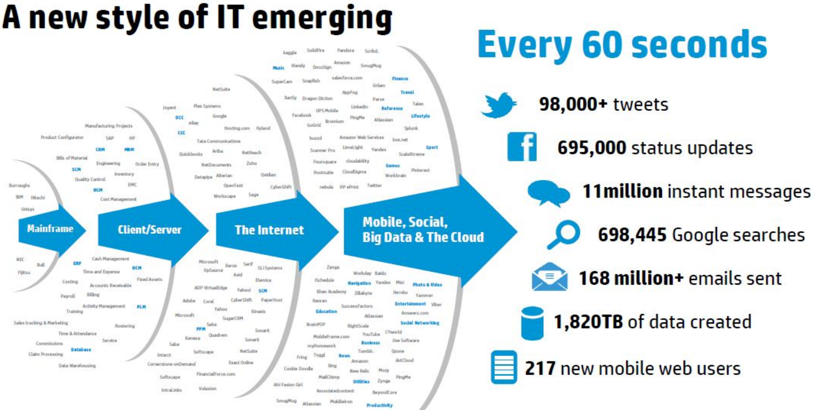


Smartphones are Spreading Everywhere

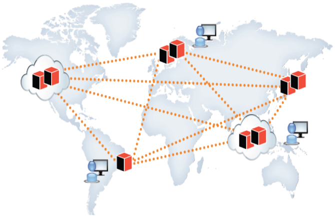


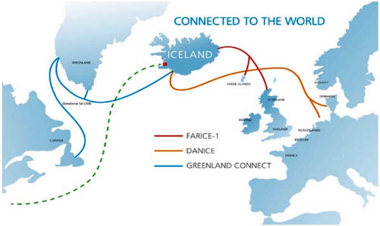


Mobile Creates 24/7 Data Deluge

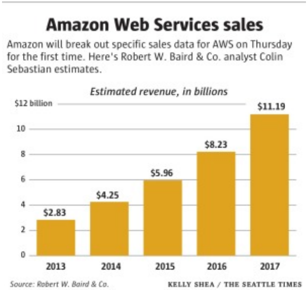


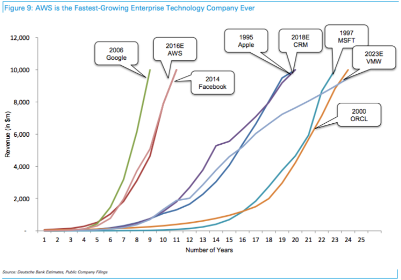
24/7 Deluge Spawns Global Computational Services





Emergence of Cloud Services





Explosion of Computational Resources

Cloud Paradigm:

- Billions of smart devices each drive queries to cloud servers
- Millions of business relying on cloud for all needs

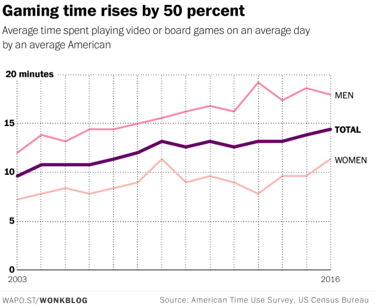
Symbiosis of cloud and economy is *lasting* and *disruptive*.

Cloud provides *any user* **same-day** delivery:

- Tens to hundreds of thousands of hours of CPU
- Pennies per CPU hour

Any user can consume *1 Million CPU hours* over a few days for a few $10K's.

Games Have *Increasing* Popularity



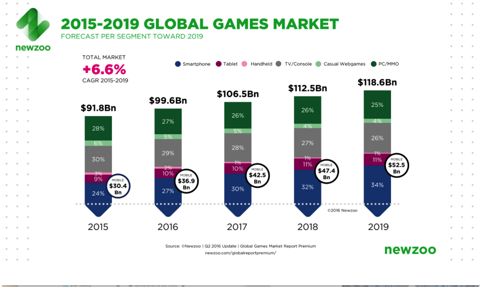
Games Have *Absurd* Popularity

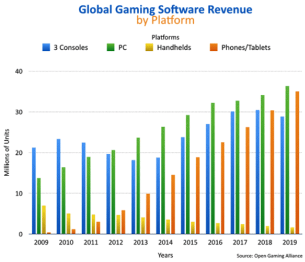
'Scientists in Antarctica are downloading mobile games. Parents in Syria worry about kids using too much tech.

JOHN KOETSIER TUNE 12/29/2015

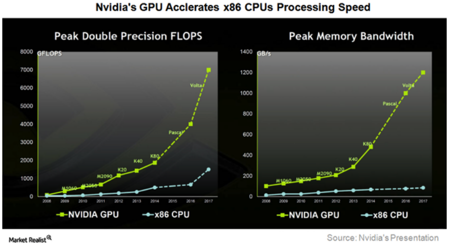
https://www.tune.com/blog/global-mobile-why-2016-is-the-global-tipping-point-for-the-mobile-economy/

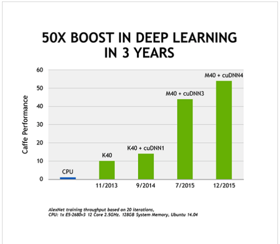
Gaming became a Massive Market





Games → GPUs → Learning Speed





Why this is happening now?

Key ingredients of DL have been in place for 25-30 years:

Landmark	Emblem	Epoch
Neocognitron	Fukushima	1980
CNN	Le Cun	mid 1980s'
Backprop	Hinton	mid 1980's
SGD	Le Cun, Bengio etc	mid 1990's
Various	Schmidhuber	mid 1980's
CTF	*DARPA etc*	*mid 1980's*

Ubiquitous massive computation now makes it possible for thousands of researchers to build, train, tear apart and rebuilt DeepNets.
Extensive Trial and Error has been necessary.

Common Task Framework (1980's)

Under CTF we have the following ingredients

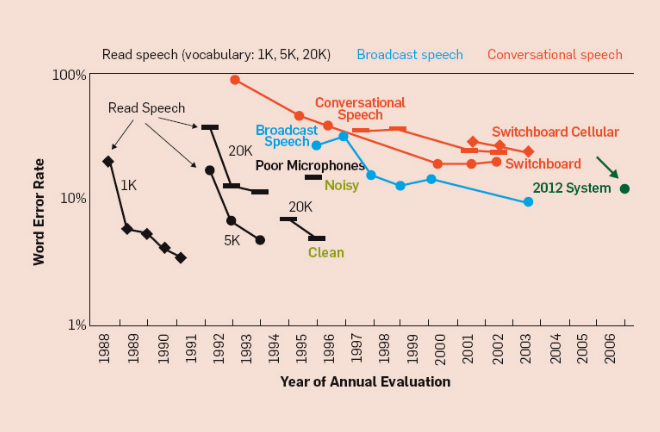
(a) A **publicly available training dataset** involving, for each observation, a list of (possibly many) feature measurements, and a class label for that observation.

(b) A set of **enrolled competitors** whose **common task** is to **infer** a class **prediction rule from the training data**.

(c) A **scoring referee**, to which competitors can submit their prediction rule. The referee runs the prediction rule against a testing dataset which is sequestered behind a Chinese wall. The referee objectively and automatically reports the score achieved by the submitted rule.

See Mark Liberman's description (Liberman, 2009).

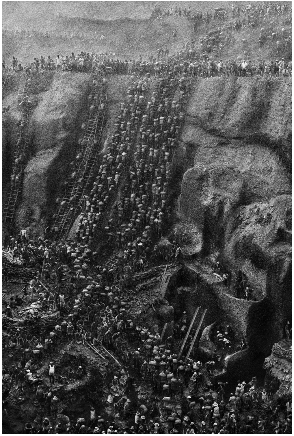
CTF *Really* Works!

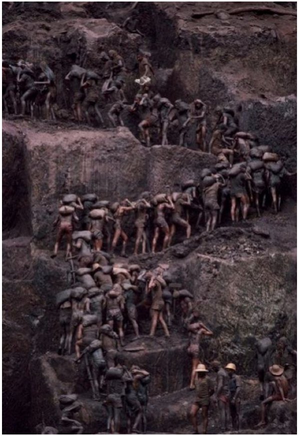


CTF Lifestyle – 1

1. Researchers set up local copies of Challenge
    - Data – Training, Test carved out of public dataset
    - Scoring – same as challenge scoring rule
2. Researcher's job: *'tuning models'*
    - Think up a family of model variations – *'tweak's*
    - Run a full *'experiment'* – suite of tweaks – *'grid'*
    - Score each tweak
    - Submit best-scoring result to central authority
3. Successful researchers perpetually motivated by *Game-ification*: tweaking, scoring, winning.
4. Researchers who tweak more often, win more often!.
5. If easier to implement tweaks and faster to evaluate them, more likely to win!.

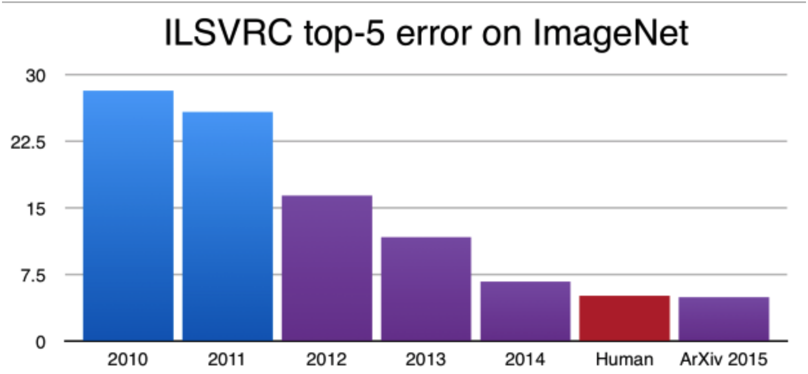
CTF Lifestyle – 2

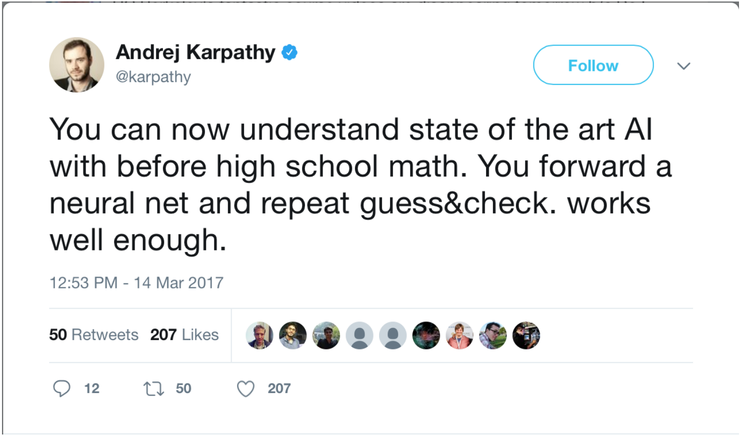




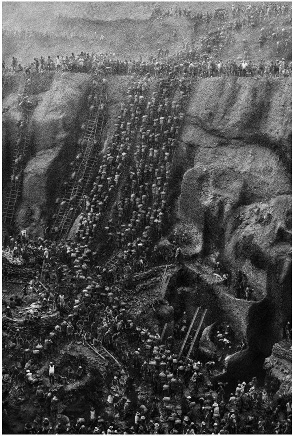
CTF Goes Mainstream

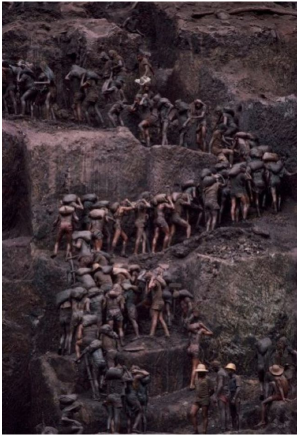
1. Netflix Challenge (2009)
   $1 Million Prize
2. Kaggle (2010)
   1 Million'th competitor expected Sept. 2017
3. Fei-Fei Li masterminds ImageNet 2008-2010
4. Hinton's Deep Learning Team wins ImageNet 2012


ILSVRC top-5 error on ImageNet
30
22.5
15
7.5
0
2010
2011
2012
2013
2014
Human
ArXiv 2015

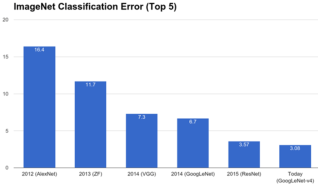

Andrej Karpathy
@karpathy
Follow
You can now understand state of the art AI with before high school math. You forward a neural net and repeat guess&check. works well enough.
12:53 PM - 14 Mar 2017
50 Retweets 207 Likes
12
50
207

Graduate Students Preparing for NIPS 2017





Global Economy → Computing → Deep Learning



SIAM NEWS MAY 2017

Research | May 01, 2017

Print

Deep, Deep Trouble

Deep Learning's Impact on Image Processing, Mathematics, and Humanity

By Michael Elad

I am really confused. I keep changing my opinion on a daily basis, and I cannot seem to settle on one solid view of this puzzle. No, I am not talking about world politics or the current U.S. president, but rather something far more critical to humankind, and more specifically to our existence and work as engineers and researchers. I am talking about...**deep learning**.

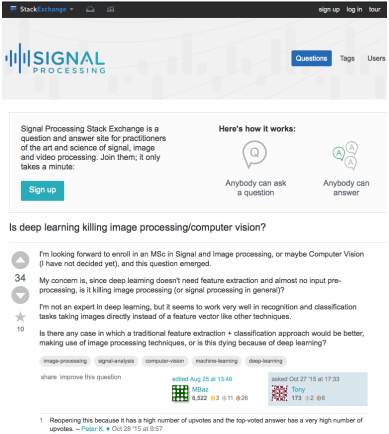
While you might find the above statement rather bombastic and overstated, deep learning indeed raises several critical questions we must address. In the following paragraphs, I hope to expose one key conflict related to the emergence of this field, which is relevant to researchers in the image processing community.

First, a few words about deep learning to put our discussion into perspective. Neural networks have been around for decades, proposing a universal learning mechanism that could, in principle, fit to any learnable data source. In its feed-forward architecture, layers of perceptrons—also referred to as neurons—first perform weighted averaging of their inputs, followed by nonlinearities such as a sigmoid or rectified-linear curves. One can train this surprisingly simple system to fit a given input set to its desired output, serving various supervised regression and classification problems.

“Deep Learning is killing intellectual life” – various

Quotes (Professors at U Wash, Princeton, MIT, ...

- “Deep Learning is killing X”, $X \in \{$ SP , IT , NLP $, \ldots \}$
- “Graduate Students only will work on deep learning”
- “One Time Extinction event” – Graduate students won’t know the fundamental tools


StackExchange sign up log in tour

SIGNAL PROCESSING Questions Tags Users

Signal Processing Stack Exchange is a question and answer site for practitioners of the art and science of signal, image and video processing. Join them; it only takes a minute:

Sign up

Here's how it works:

Anybody can ask a question

Anybody can answer

Is deep learning killing image processing/computer vision?

34

10

I'm looking forward to enroll in an MSc in Signal and Image processing, or maybe Computer Vision (I have not decided yet), and this question emerged.

My concern is, since deep learning doesn't need feature extraction and almost no input pre-processing, is it killing image processing (or signal processing in general)?

I'm not an expert in deep learning, but it seems to work very well in recognition and classification tasks taking images directly instead of a feature vector like other techniques.

Is there any case in which a traditional feature extraction + classification approach would be better, making use of image processing techniques, or is this dying because of deep learning?

image-processing signal-analysis computer-vision machine-learning deep-learning

share improve this question

edited Aug 25 at 13:46
MBaz
6,522 3 11 26

asked Oct 27 '15 at 17:33
Tony
173 2 6

1 Reopening this because it has a high number of upvotes and the top-voted answer has a very high number of upvotes. – Peter K. ♦ Oct 28 '15 at 9:57

D!gitalist Magazine

19-July-2017 | Digital Economy | Hyperconnectivity

Are AI And Machine Learning Killing Analytics As We Know It?

Joerg Koesters

According to IDC, artificial intelligence (AI) is expected to become pervasive across customer journeys, supply networks, merchandizing, and marketing and commerce because it provides better insights to optimize retail execution. For example, in the next two years:

- 40% of digital transformation initiatives will be supported by cognitive computing and AI capabilities to provide critical, on-time insights for new operating and monetization models.
- 30% of major retailers will adopt a retail omnichannel commerce platform that integrates a data analytics layer that centrally orchestrates omnichannel capabilities.

One thing is clear: new analytic technologies are expected to radically change analytics –

Intelligent Machines

The Dark Secret at the Heart of AI

No one really knows how the most advanced algorithms do what they do. That could be a problem.

by Will Knight   April 11, 2017

Last year, a strange self-driving car was released onto the quiet roads of Monmouth County, New Jersey. The experimental vehicle, developed by researchers at the chip maker Nvidia, didn't look different from other autonomous cars, but it was unlike anything demonstrated by Google, Tesla, or General Motors, and it showed the rising power of artificial intelligence. The car didn't follow a single instruction provided by an engineer or programmer. Instead, it relied entirely on an algorithm that had taught itself to drive by watching a human do it.

AI Learns Racism/Sexism

See:

https://www.theguardian.com/technology/2017/aug/10/faceapp-forced-to-pull-racist-filters-digital-blackface

http://www.independent.co.uk/life-style/gadgets-and-tech/news/ai-robots-artificial-intelligence-racism-se

Where are the Intellectuals?

Common reactions...

- Meh....
- "This is not happening"
- "This is a Crisis"
- "This is what I've been telling you for years"

Theory in Crisis

Some Theory “Lessons”:

- “There’s no magic method”
- “Curse of Dimensionality”
  - ... Approximation Theory
  - ... Statistical Modelling
  - ... Optimization Theory

If those are the “Lessons”, Theory Fails

Recent “Student Insurrections”

Theorists are responding

That's this course!

Field	Example
Neuroscience	Bruno Olshausen
Harmonic Analysis	Joan Bruna and Stephane Mallat Helmut Boelcskei and co-authors Vardan Papyan, Jeremias Sulam, Yaniv Romano and Michael Elad
Approximation Theory	Tomaso Poggio and Hrushikesh Mhaskar
Statistics	Zaid Harchaoui
Information Theory	Naftali Tishby

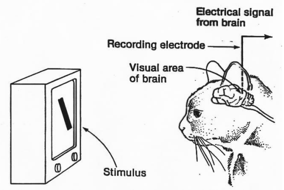
Deep Learning as a Magic Mirror

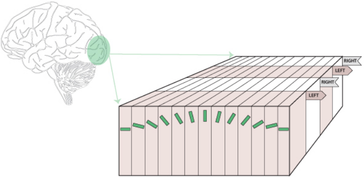
Figure : Every theorist who looks at it see what they wish

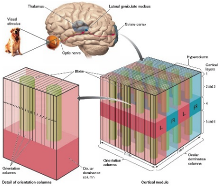
Goal of Theory

- What is a Theory?
  - Vulgar meaning – any model, any set of formal arguments.
  - Precise meaning – models that explain and that predict.
- What can Theory contribute?
  - Analysis
  - Prediction
- Should/Can there be Theories of Deep Learning?

Visual Neuroscience – Hubel/Wiesel et seq.

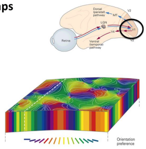






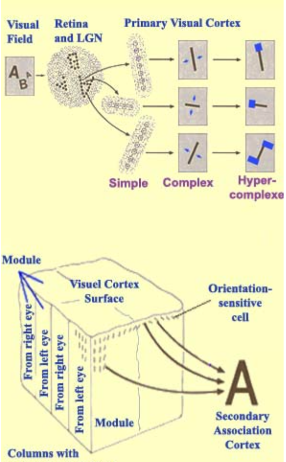
V1 orientation maps

- Continuous maps of orientation preference: "pinwheels"
- Consistent preference though the depth of cortex: columnar architecture



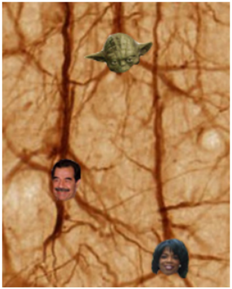
Kandell et. al., *Princ. of Neural Sci.*

Simple Cells/Complex Cells



Grandmother Cells

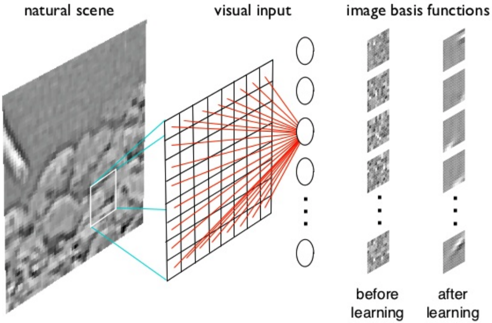
Four years ago, Rodrigo Quian Quiroga from Leicester University showed that single neurons in the brain react selectively to the faces of specific people, including celebrities like Halle Berry, Jennifer Aniston and Bill Clinton. Now, he's back, describing single neurons that respond selectively to the concept of Saddam Hussein or Oprah Winfrey. This time, Quiroga has found that these neurons work across different senses, firing to images of Oprah or Saddam as well as their written and spoken names.



In one of his volunteers, Quiroga even found a neuron that selectively responded to photos of himself! Before the study began, he had never met the volunteers in the study, which shows that these representations form very quickly, at least within a day or so.

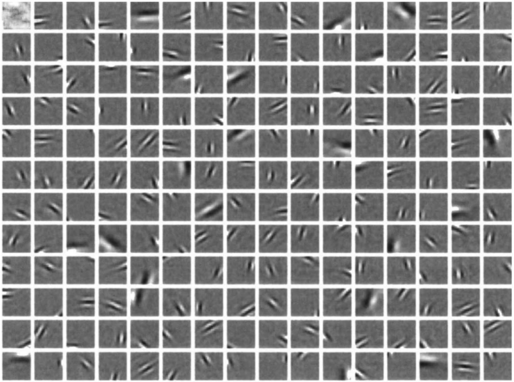
Olshausen and Field (1996)

Efficient coding of natural images: Olshausen and Field, 1996



Network weights are adapted to maximize coding efficiency:
minimizes redundancy and maximizes the independence of the outputs

Olshausen and Field



Prediction/Inspiration by Neuroscience

Experimental Neuroscience uncovered the

- ... neural architecture of Retina/LGN/V1/V2/V3/ etc
- ... existence of neurons with weights and activation functions (simple cells)
- ... pooling neurons (complex cells)

All these features are somehow present in today's sucessful Deep Learning systems

Neuroscience	Deep Network
Simple cells	First layer
Complex celle	Pooling Layer
Grandmother cells	Last layer

Theorists Olshausen and Field (Nature, 1996) demonstrated that receptive fields learned from image patches

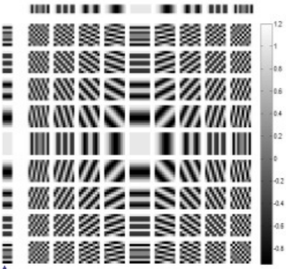
- Olshausen and Field showed that optimization process can drive learning image representations
- Olshausen-Field representations bear strong resemblance to defined mathematical objects from harmonic analysis wavelets, ridgelets, curvelets
- Harmonic analysis: long history of developing optimal representations via optimization
- Research in 1990's: Wavelets etc are optimal sparsifying transforms for certain classes of images
- Relevant Talks:

Speaker	Institution	Date
Helmut Boelsckei	ETHZ	October 11
Joan Bruna	Courant/NY	) Novermber 15

Optimal Representations – eigenfunctions – fourier

real $e^{-j2\pi(\frac{um}{N}+\frac{vn}{N})}$

imag $e^{-j2\pi(\frac{um}{N}+\frac{vn}{N})}$

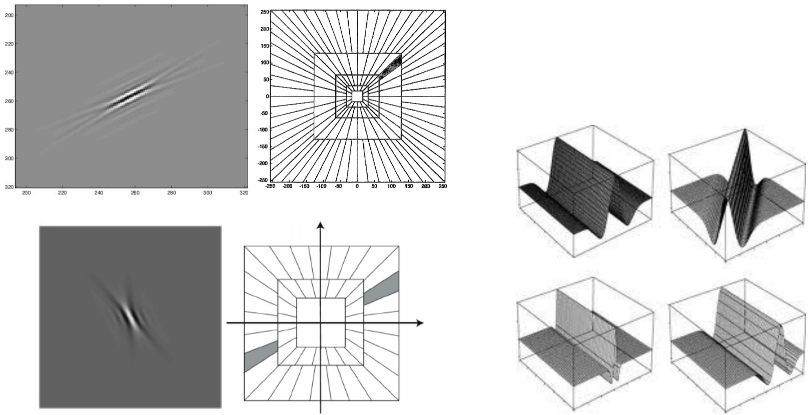


real($e^{-j2\pi\frac{um}{M}}$)

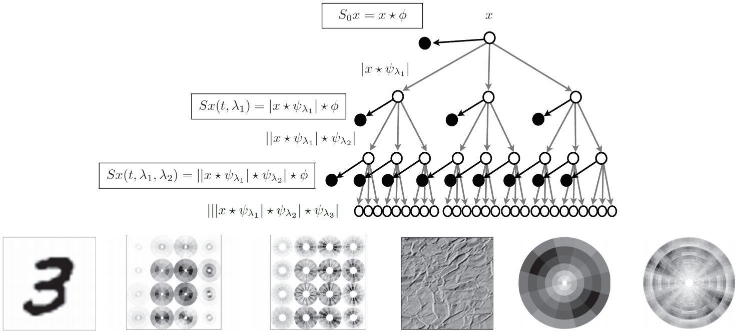
imag($e^{-j2\pi\frac{um}{M}}$)

Multiscale Representations
Wavelets – Almost Eigenfunctions

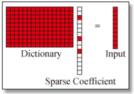
Geometric Multiscale – Ridgelets, Curvelets and Shearlets

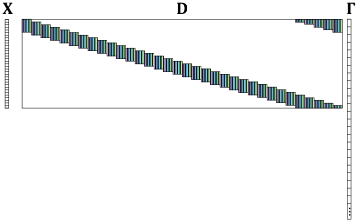


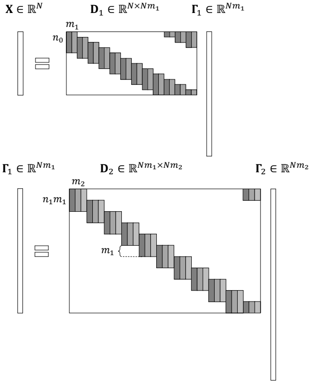
Scattering Transform; Mallat (2012), Bruna and Mallat (2014)



Sparse Representations







Approximation Theory, I

- Class prediction rule can be viewed as function $f(x)$ of high-dimensional argument
- *Curse of Dimensionality*
  - Traditional theoretical obstacle to high-dimensional approximation
  - "*Functions of high dimensional x can wiggle in too many dimensions to be learned from finite datasets*"

Approximation Theory, II

- Ridge Functions $\rho(u'x)$ mathematically same as deep learning first layer outputs.
- Sums of Ridge Functions mathematically same as input to second layer.
- Approximation by Sums of Ridge Functions $f \approx \sum_i \rho_i(u_i'x)$ studied for decades
- Theorists (1990's-Today): certain functions $f(x)$ approximated by ridge sums with no curse of dimensionalty

Approximation Theory, III

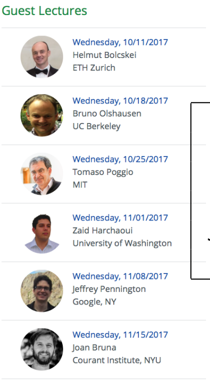
- Compositional functions $f(x) = h(g_1(x_{i_{1,1}}, ..., x_{i_{1,k}}), g_2(x_{i_{2,1}}, \dots, x_{i_{2,k}}), \dots, g_\ell(x_{i_{\ell,1}}, \dots, x_{i_{\ell,k}}))$ are functions of small number of functions; $\ell, k \ll d$.
- VGG Nets are deep compositions
- Approximation by Compositional Functions studied for decades
- Theorists (1990's-Today): certain functions $f(x)$ avoid curse of dimensionalty using multilayer compositions
- T. Poggio (MIT) and Hrushikesh Mhaskar (Caltech) have several papers analyzing deepnets as deep compositions.
- Poggio will speak to us October 25.

Modern Statistical Theory

Interactions with key themes in modern statistical theory

- Estimation of functions in high dimensions
- VC Classes/Generalization Bound
- Statistical Models
  - Projection Pursuit
  - Compositional Models
  - Generalized Additive Models
- Regularization/Overfitting/Model Selection
- Experimental Design
- Observational vs Experimental Data
- Ecological Correlation/Correlation vs Causation

A Look Ahead: `https://stats385.github.io`



Helmut Boelcskei	ETH Zuerich
Bruno Olshausen	Redwood Center for Theoretical Neuroscience
Tomaso Poggio	MIT
Zaid Harchaoui	Univ Wash Seattle
Jeffrey Pennington	Google
Joan Bruna	NYU

Course Structure

1. Review DL concepts
2. Review Theoretical Approaches
3. Specific Theoretical Contributions – Mostly guest lectures
4. Postmortem, last 3 lectures.

Global Economy → Computing → Deep Learning