Data Science Training

November 2017

Introduction to Deep Learning

Xavier Bresson

Data Science and AI Research Centre
NTU, Singapore

http://data-science-optum17.tk

Outline

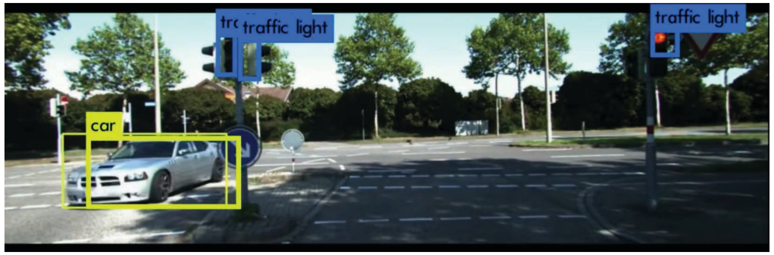
- History of Deep Learning
- What makes DL great?
- DL architectures
- Future of DL
- How to develop AI products?
- How to develop skills/career in AI?
- Conclusion

Outline

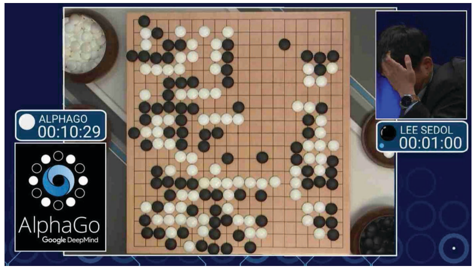
- History of Deep Learning
- What makes DL great?
- DL architectures
- Future of DL
- How to develop AI products?
- How to develop skills/career in AI?
- Conclusion

The Deep Learning Revolution

- Breakthrough in the last 5 years on some AI tasks:



Self-driving cars



Go game

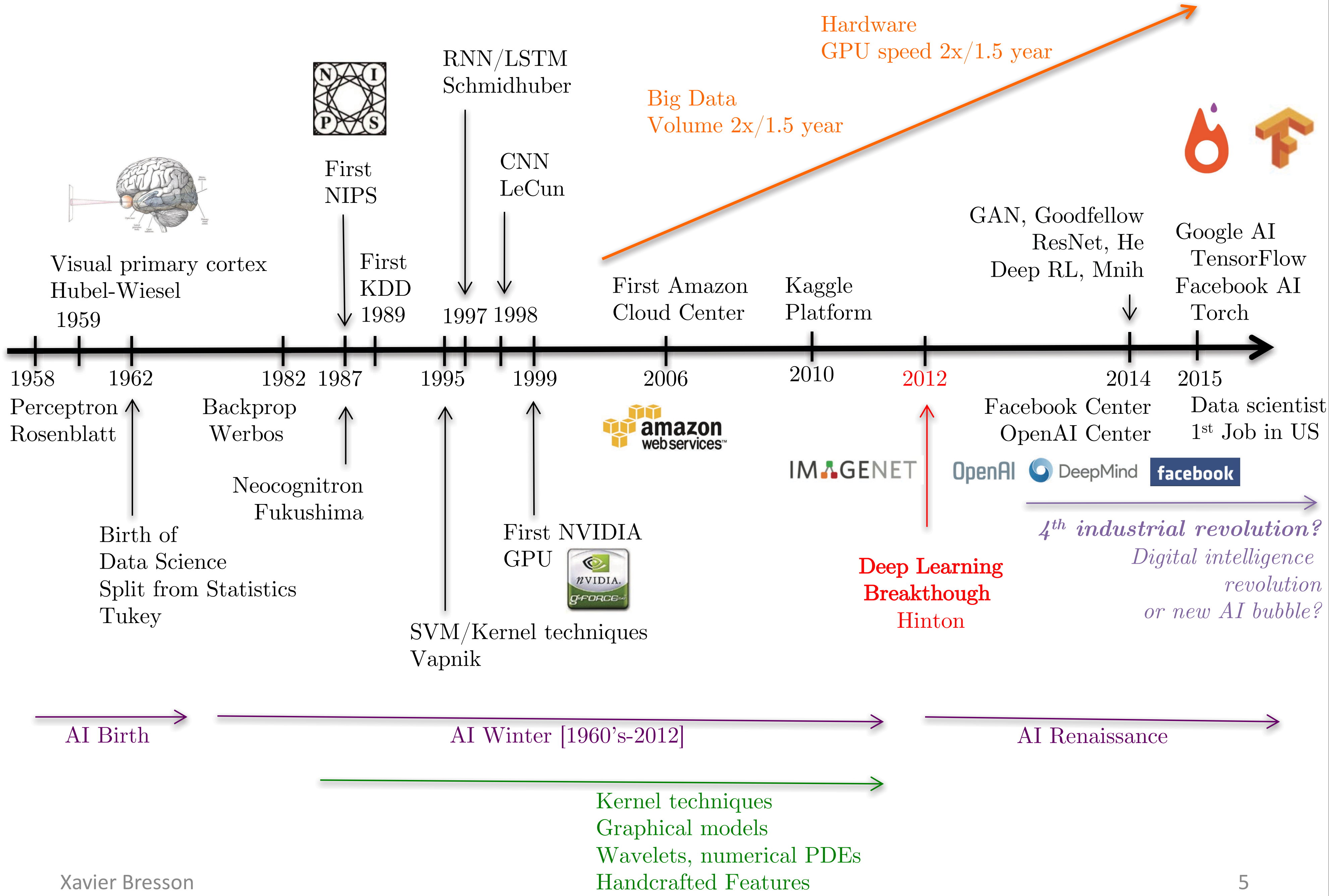
Input sentence:	Translation (PBMT):	Translation (GNMT):	Translation (human):
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。	Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang will start the annual dialogue mechanism with Prime Minister Trudeau of Canada and hold the first annual dialogue between the two premiers.	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.

Machine translation

Speech recognition

- At the core, artificial neural networks a.k.a. deep learning

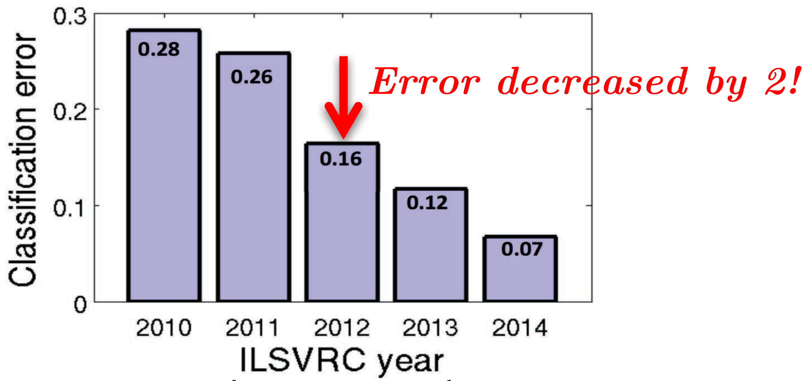
A Brief History of Deep Learning



2012: Deep Learning Breakthrough

- **ImageNet [Fei Fei-et.al.'09]:** International Image Classification Challenge
  1,000 object classes and 1,431,167 images

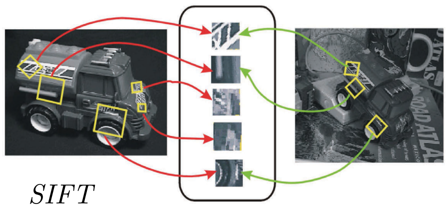
[L. Fei-Fei-et.al.'09]



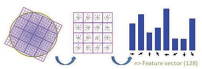
[Russakovsky-et.al.'14]

- **Before 2012:** Handcrafted filters + SVM classification

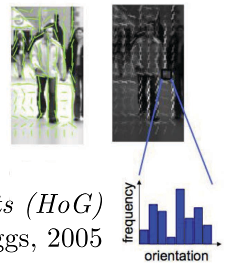
[Fei-Fei Li-et.al.'15]

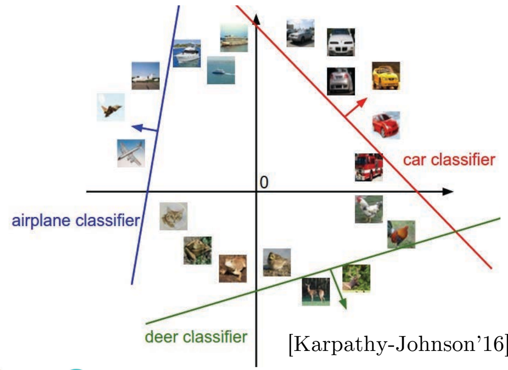


SIFT
Lowe, 1999
(most cited paper in CV)



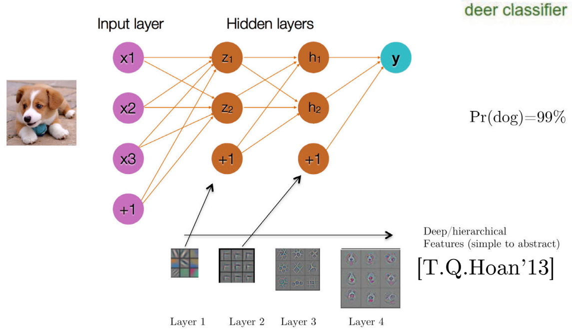
Histogram of Gradients (HoG)
Dalal & Triggs, 2005





[Karpathy-Johnson'16]

After 2012: *Learn filters and classifier with neural networks (end-to-end system).*



[T.Q.Hoan'13]

Outline

- History of Deep Learning
- **What makes DL great?**
- DL architectures
- Future of DL
- How to develop AI products?
- How to develop skills/career in AI?
- Conclusion

What are the Principles behind Deep Learning?

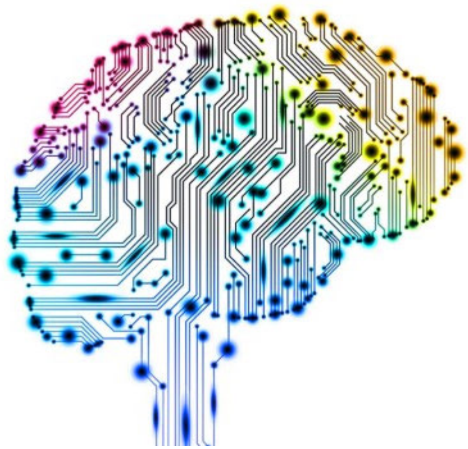
Biological systems

Principles of Aerodynamics

Artificial systems

?

Principles of Intelligence ***inspired*** *by the brain*



Principle 1: End-to-End Learning

- Brain is an end-to-end system:

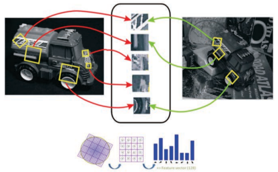
$$\textit{Raw data} \Rightarrow \textit{Solution of the given task}$$

- Vision (visual cortex neurons): *Light* $\Rightarrow$ *Object recognition*

Fixed/Handcrafted Feature Extractor [1963-2012]:

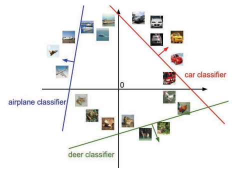
hand-crafted Feature Extractor → "Simple" Trainable Classifier → *Car*

Traditional Learning



SIFT
Lowe, 1999

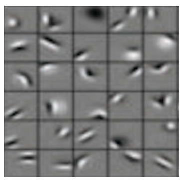
[LeCun.'17]

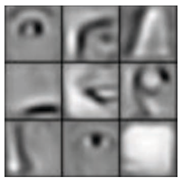


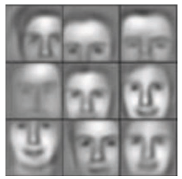
Learned Feature Extractor + Classifier [2012-]:

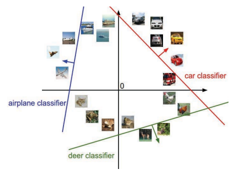
Low-Level Features → Mid-Level Features → High-Level Features → Trainable Classifier → *Car*

End-to-end Learning



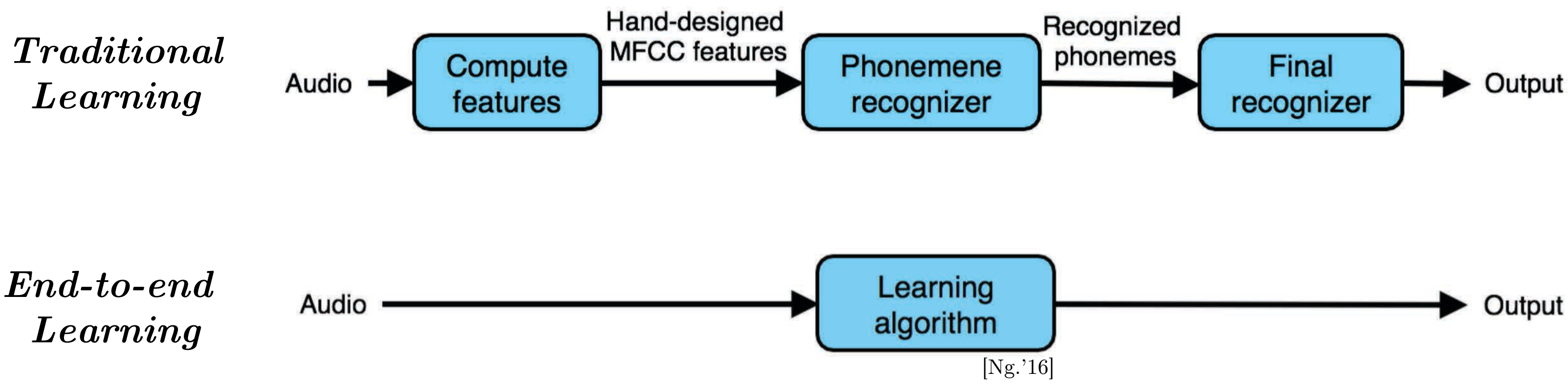






Principle 1: End-to-End Learning

- **Hearing (auditory neurons):** ***Sound waves*** $\Rightarrow$ ***Word recognition***



- **Lesson: Never** (or as less as possible) handcraft any feature detector!

- Deep learning revolution is about ***supervised*** end-to-end learning.

Supervised Learning

- **All successful DL systems are supervised, i.e. they use labeled data.**
  For each input data (image, audio, text, etc), the output solution is known (object class, spoken word, sentiment, etc).

Problem	X	Y
Spam classification	Email	Spam/Not spam (0/1)
Image recognition	Image	Integer label
Housing price prediction	Features of house	Price in dollars
Product recommendation	Product & user features	Chance of purchase

Problem	X	Y
Image captioning	Image	Text
Machine translation	English text	French text
Question answering	(Text,Question) pair	Answer text
Speech recognition	Audio	Transcription
TTS	Text features	Audio

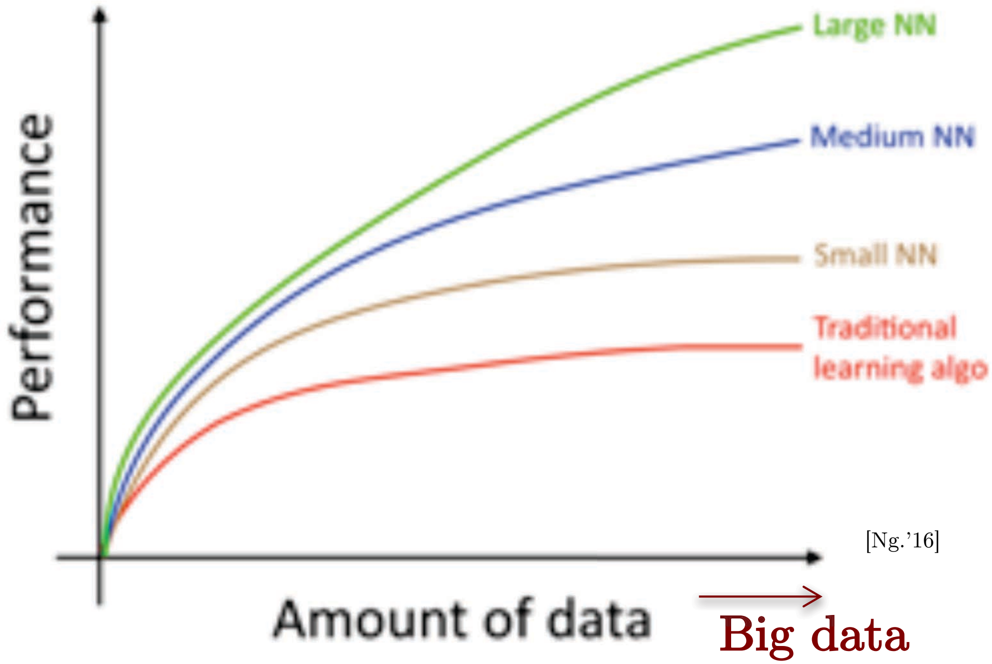
[Ng.'16]

Principle 2: Data Drives Learning

- **Human brain receives 2.5 petabytes of information in a lifetime.**
  $\Rightarrow$ ***Large-scale/big data is essential for smart learning systems.***

- **Only deep learning systems are able to absorb big data:**

Data and Machine Learning



[Ng.'16]

Labeled Data vs Unlabeled Data

3rd industrial revolution

- **We live in a data society:** Huge quantity of data available ☺, but most data are not directly useful for deep learning ☹.

- **Supervised** learning require (lots of) **labeled** data, but:
  1. Labeling is ***time consuming*** (each data to a class).
  2. It demands human brain (learning ability ***bounded*** by human intelligence).
  3. Some important data is ***not accessible*** (e.g. medical data, nuclear meltdown).

- **Most data are unlabeled** ⇒ **Unsupervised** learning, open problem

- One of the biggest challenges: Collect labeled data. **Systems are as smart as the data they use!** Ex. Google self-driving cars collected 2 millions of miles, is it enough to guarantee super-human performances?

Principle 3: Deep Architectures

- Human brain is composed of **cascades of neuron layers.**

- **Hierarchical layers are able to learn high-dimensional data** because they can represent exponential functions (and beat the curse of dimentionality).

Data dimensions:

dim(Go) $=$ 19 x 19 ≈ 10^3

dim(Images) $=$ 512 x 512 ≈ 10^6

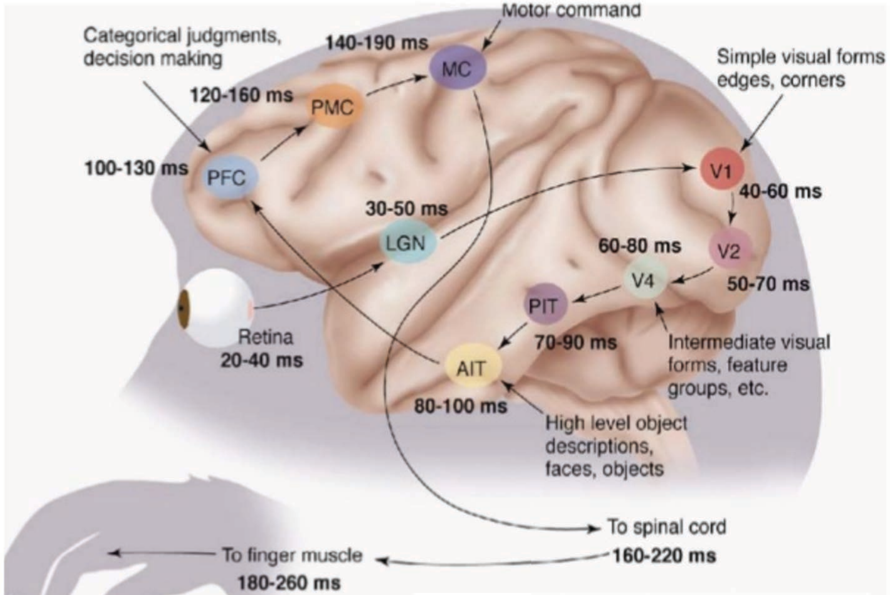
dim(Books) $=$ 80,000 ≈ 10^5

All data are intractable:

#Go $= 3^{19 \text{ x } 19} \approx 10^{172}$

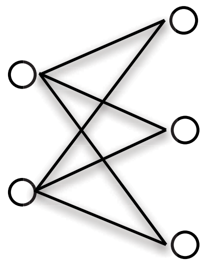
#Images $= 256^{512 \text{ x } 512} \approx 10^{630,000}$

#Books $= 8{,}000^{80,000} \approx 10^{310,000}$



[Thorpe.'01]

$n\_in = 2$



$n\_out = 3$

#configurations= $2^{n\_out} = 2^3 = 8$

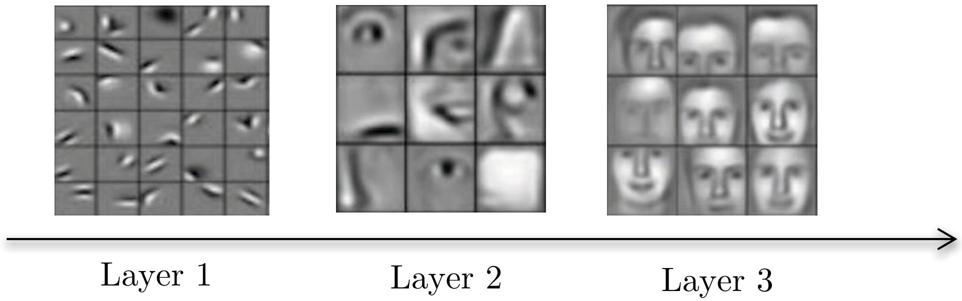
Parameters

|W|= 2.3

⇒ Number of parameters grow **linearly**, while function capacity grows **exponentially**.

Data Compositionality

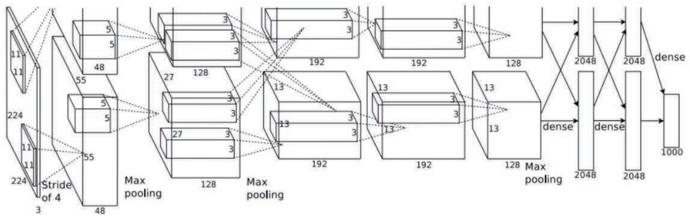
- **Real Data are compositional (special case of exponential representations):** They are composed of simple structures that agglomerate to compose a little more complex and abstract structures, and so on.



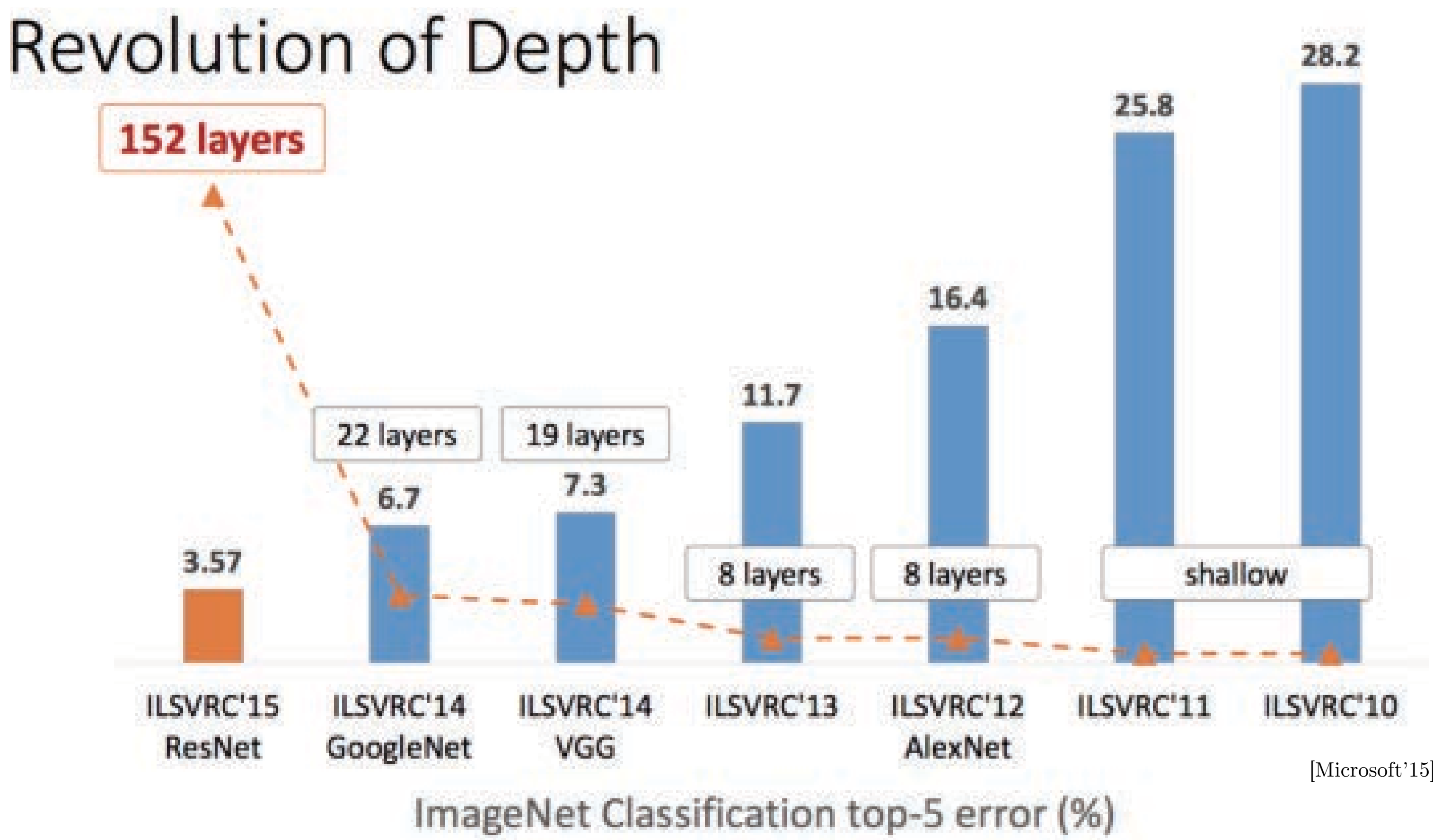
Simple to abstract structures

- **Deep architectures can capture the compositionality of data.**

2012 DL
***7-layer** NN*
[Krizhevsky-et.al]

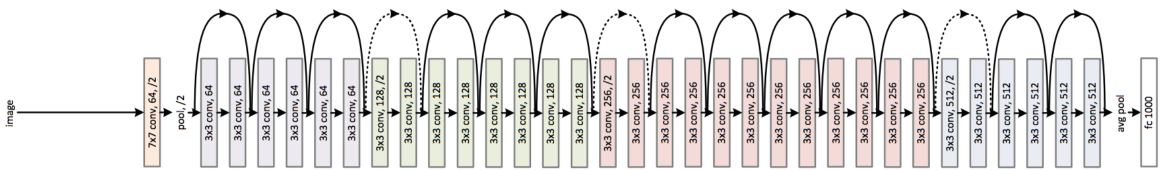


Inflation of Deep Architectures



[Microsoft'15]

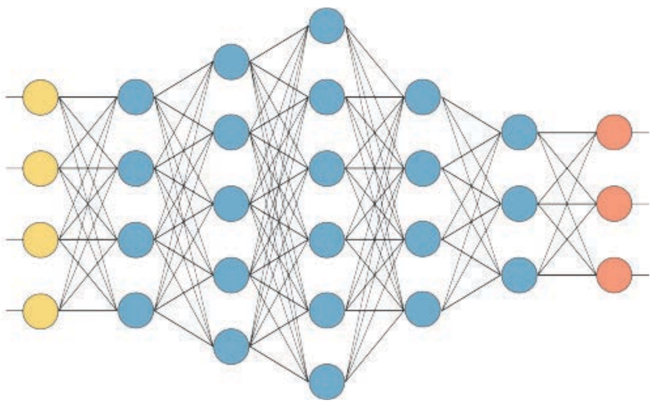
2015 DL
ResNet ***152-layer*** *NN*
[He-et.al]



Principle 4: Weight Learning

- **Brain has a mechanism to change its neuron strengths to learn any task.** Ex: Learning to drive, to speak a new language, etc.

- [Hinton'12] At 1st order, this mechanism can be approximated as a **stochastic gradient descent algorithm** called **backpropagation**.



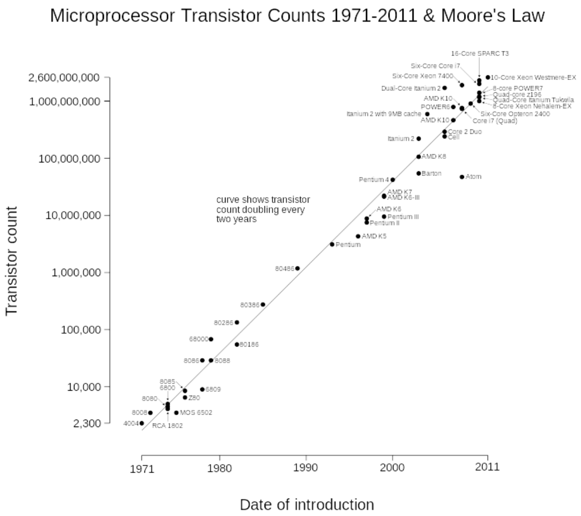
- SGD is a sound mathematical technique that is guaranteed to work
  $\Rightarrow$ It is possible to learn **NN architectures with billions of parameters**!

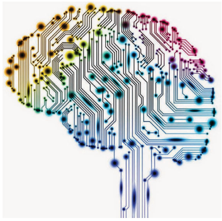
- Issue: **SGD is a very slow process** (CFL condition).

Deep Learning Infrastructure

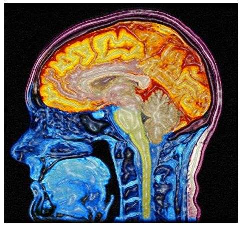
- Rise of fast computational power (CPU, GPU, Cloud computing) thanks to **Moore's law**: 100x faster every decade.

  Note: Brain 10^{14} synapses/10^{3} Tflops = Computers in 2030 (100 times faster than today)





2030
=



- Deep learning computations: Mostly **matrix multiplications** (Ax)
  $\Rightarrow$ **Highly parallelizable** on GPUs and Cloud computing.

Nvidia Titan X: 11 Tflops

Deep Learning Infrastructure

Implementation frameworks

Top 3 softwares
Nov 2017

Keras

Deep-able hardware

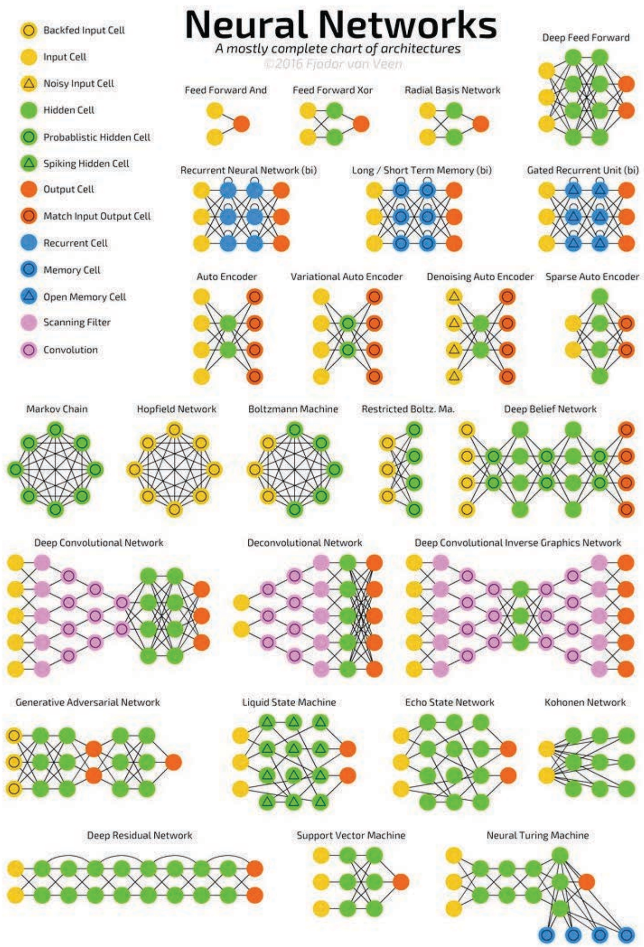
Outline

- History of Deep Learning
- What makes DL great?
- **DL architectures**
- Future of DL
- How to develop AI products?
- How to develop skills/career in AI?
- Conclusion

Deep Learning Architectures

- **Most successful architectures:**
  - **Convolutional Neural Networks:** Computer Vision, Video
  - **Recurrent Neural Networks:** NLP, Machine translation, speech-to-text
  - **Deep Reinforcement Learning:** AlphaGo, Robotics

- **Recent progresses:**
  - Memory Networks
  - Attention Networks
  - Generative Adversarial Networks
  - Learn-to-learn Networks
  - Transfer learning

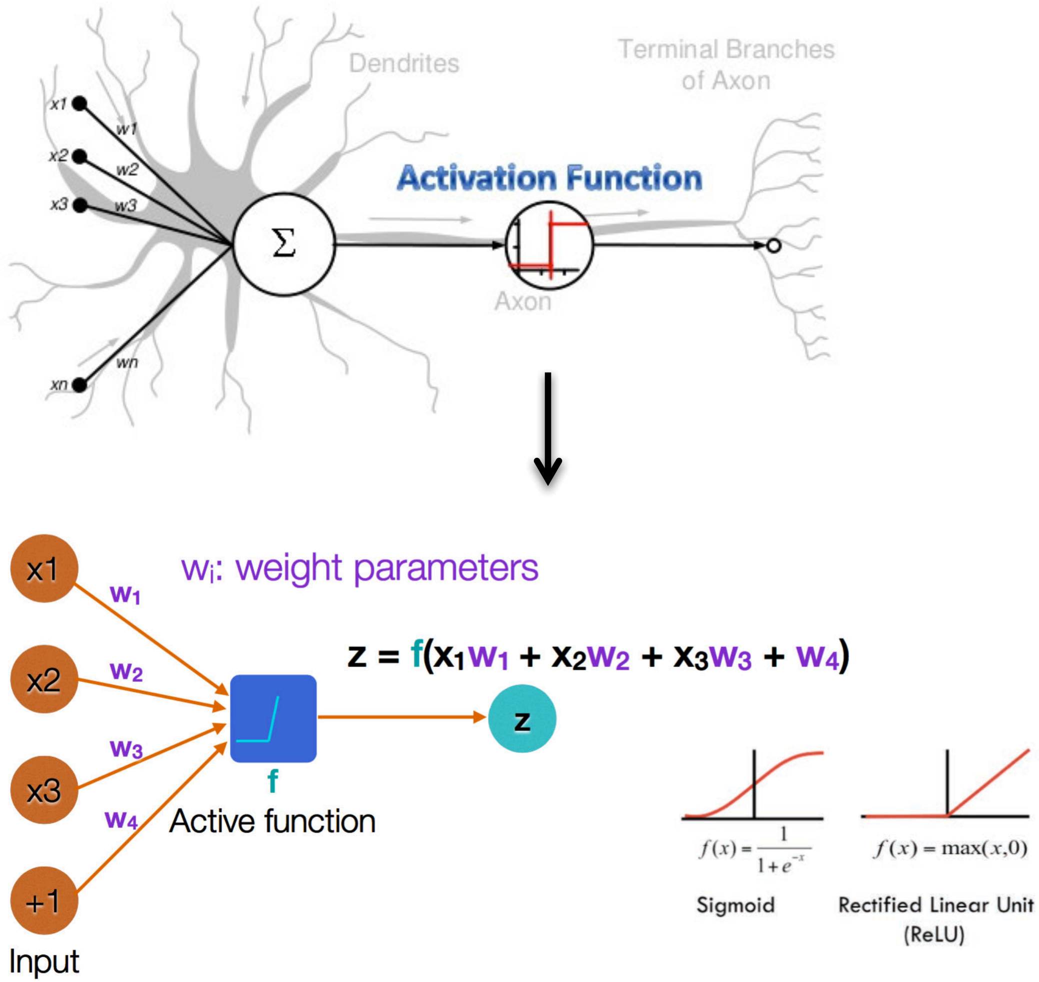


Simplest Deep Learning Architecture

- Basic neuron (1 layer):

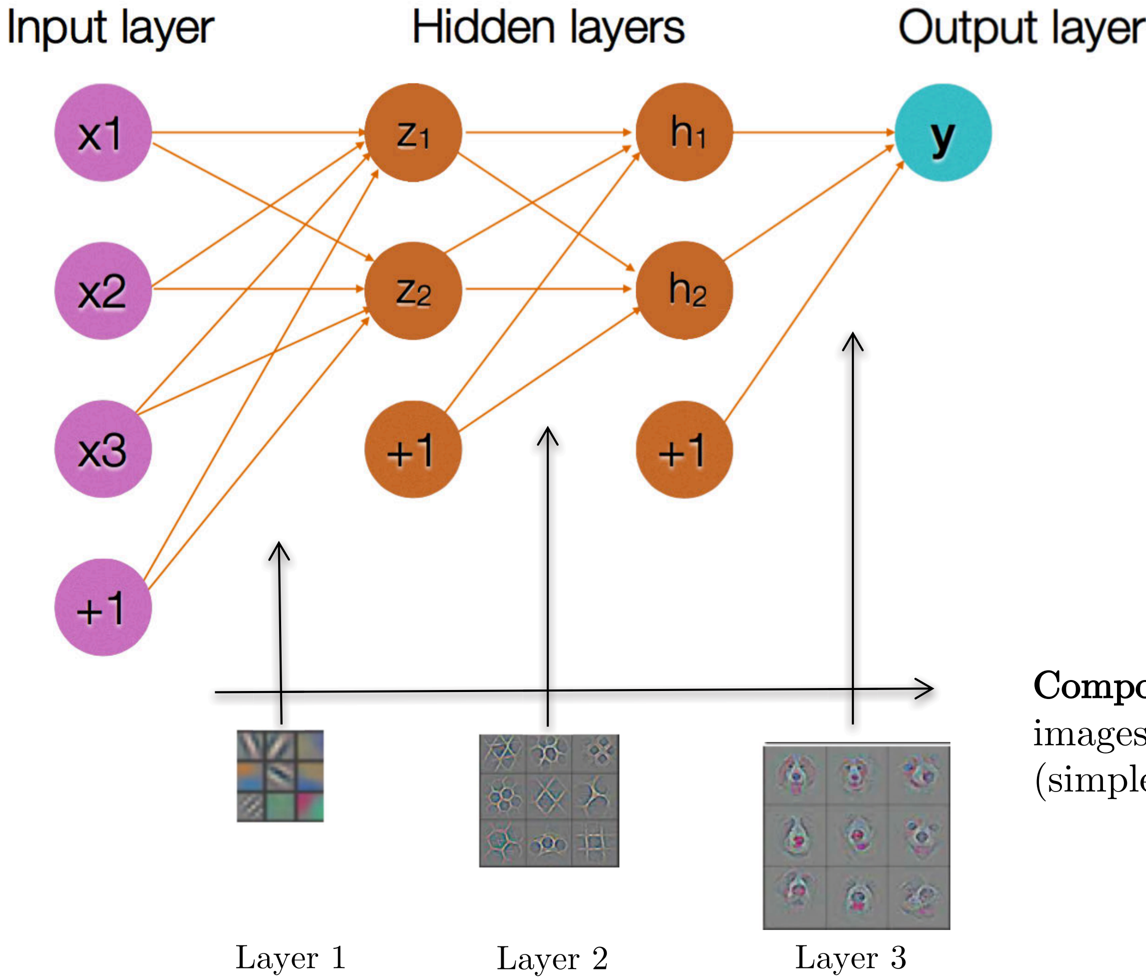
$$z=f(\underbrace{Wx+b}_{\text{Linear}})$$

Non-linear



- Then, we can cascade multiple layers.

Simplest Deep Learning Architecture



$Pr(\text{dog}) = 99\%$

Compositionality in images: Deep features (simple to abstract)

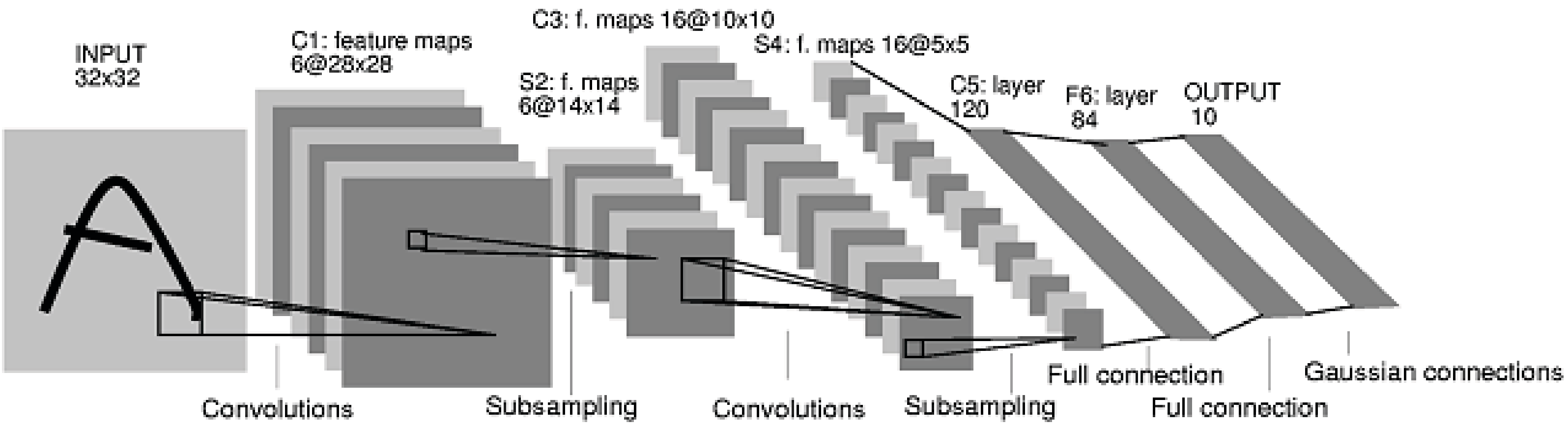
- *Note:* Brain has 100 billion neurons and 1 trillon axons.

Convolutional Neural Networks [LeCun-et.al'98]

- State-of-the-art NNs for **Computer Vision** problems, inspired by **human visual primary cortex**.

- Important principle of CNN: **shift-invariance** recognition.

Yann LeCun
(Facebook, NYU)



[LeCun.'98]

- **Breakthrough** in 2012 with ImageNet challenge.

Application of CNN

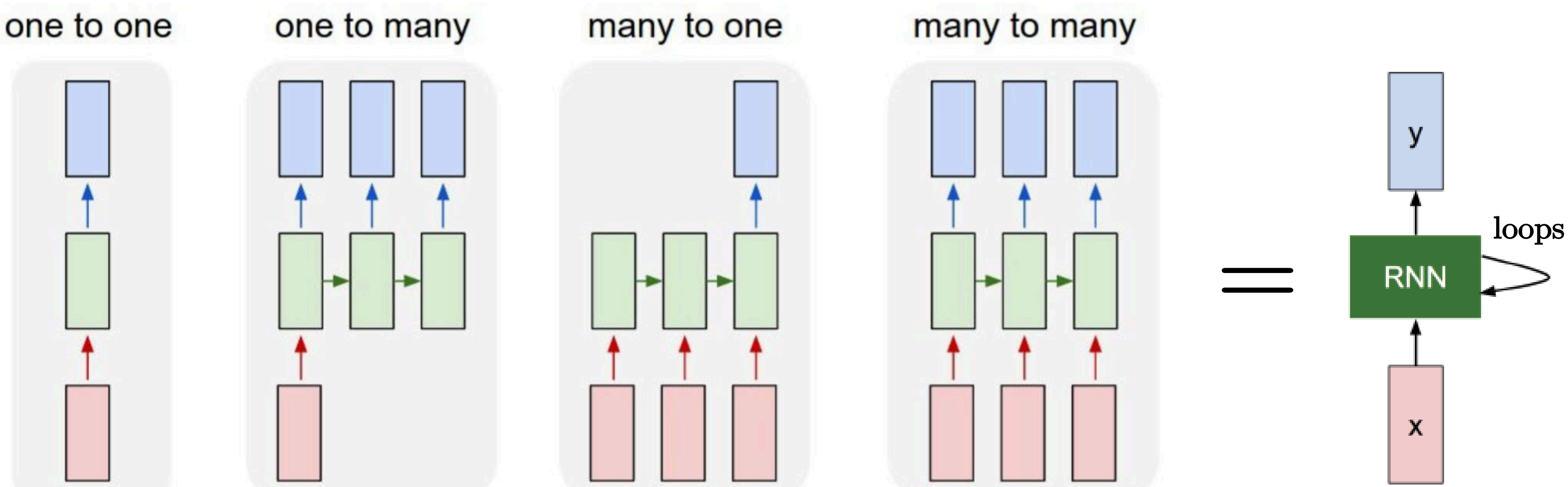
- Self-driving cars

Recurrent Neural Networks [Schmidhuber-et.al'97]

- State-of-the-art NNs for **sequence of data** like time series, natural language processing, speech-to-text, text analysis, etc.

Juergen Schmidhuber
(USI, Switzerland)

- Recurrent Networks offer a lot of **flexibility**:



[Karpathy.'15]

Image Captioning
Image ⇒
Sequence of words

Sentiment Classification
Seq of words ⇒
Sentiment

Machine Translation
Seq of words ⇒
Seq of words

Recurrence formula:

$$\boxed{h_t} = \boxed{f_W}(\boxed{h_{t-1}}, \boxed{x_t})$$

new state
old state
input vector at some time step
some function with parameters W

Application of RNN

- Machine translation (NLP)

Google's Neural Machine Translation system (Sept 27, 2016)

"Today we announce the Google Neural Machine Translation system (GNMT), which utilizes state-of-the-art training techniques to achieve **the largest improvements to date for machine translation quality**."

Input sentence:	*Translation (PBMT):*	*Translation (GNMT):*	*Translation (human):*
李克強此行將啟動中加總理年度對話機制，與加拿大總理杜魯多舉行兩國總理首次年度對話。	Li Keqiang premier added this line to start the annual dialogue mechanism with the Canadian Prime Minister Trudeau two prime ministers held its first annual session.	Li Keqiang will start the annual dialogue mechanism with Prime Minister Trudeau of Canada and hold the first annual dialogue between the two premiers.	Li Keqiang will initiate the annual dialogue mechanism between premiers of China and Canada during this visit, and hold the first annual dialogue with Premier Trudeau of Canada.

Application of RNN

- **Text generation:** Learn to write like Shakespeare.

Sonnet 116 – Let me not ...

by William Shakespeare

Let me not to the marriage of true minds
Admit impediments. Love is not love
Which alters when it alteration finds,
Or bends with the remover to remove:
O no! it is an ever-fixed mark
That looks on tempests and is never shaken;
It is the star to every wandering bark,
Whose worth's unknown, although his height be taken.
Love's not Time's fool, though rosy lips and cheeks
Within his bending sickle's compass come:
Love alters not with his brief hours and weeks,
But bears it out even to the edge of doom.
If this be error and upon me proved,
I never writ, nor no man ever loved.

at first:

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

train more

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

train more

```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

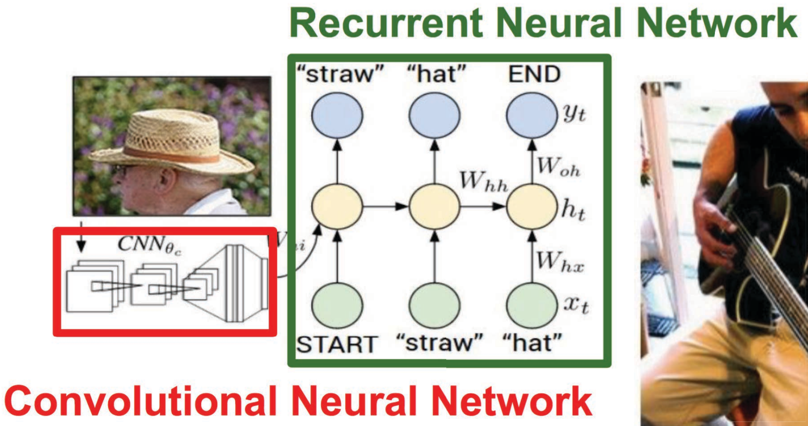
train more

```
"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
```

[Karpathy.'15]

Application of RNN

- Image Captioning




"man in black shirt is playing guitar."


"construction worker in orange safety vest is working on road."


"two young girls are playing with lego toy."


"a young boy is holding a baseball bat."

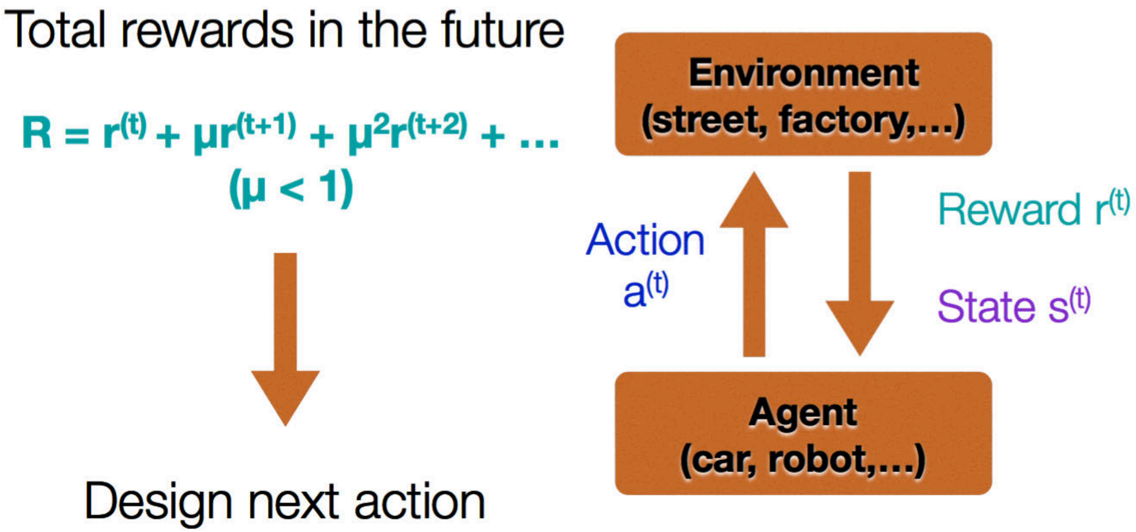

"a cat is sitting on a couch with a remote control."


"a woman holding a teddy bear in front of a mirror."

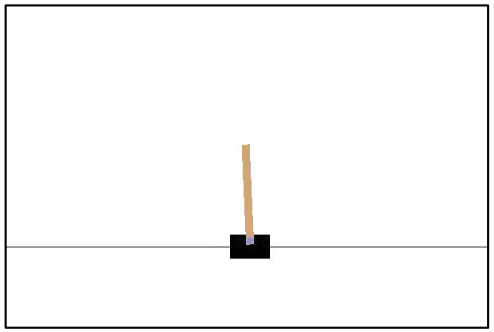
[Karpathy.'15]

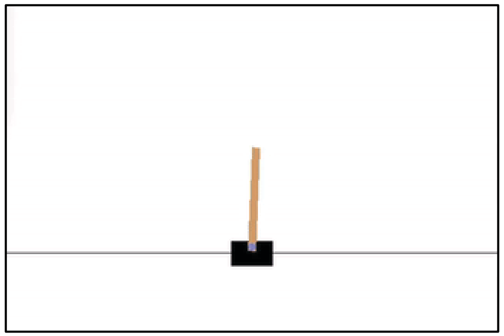
Deep Reinforcement Learning [DeepMind'15]

- **Machine can learn by itself:**



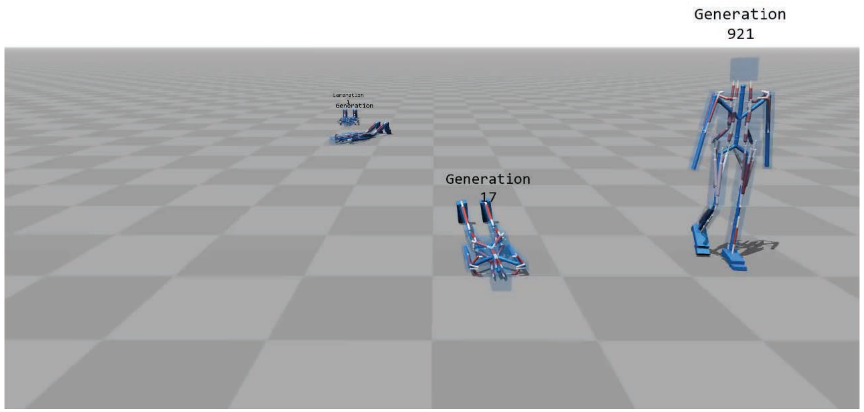
- **Ex:** Cart-Pole (Inverted Pendulum)

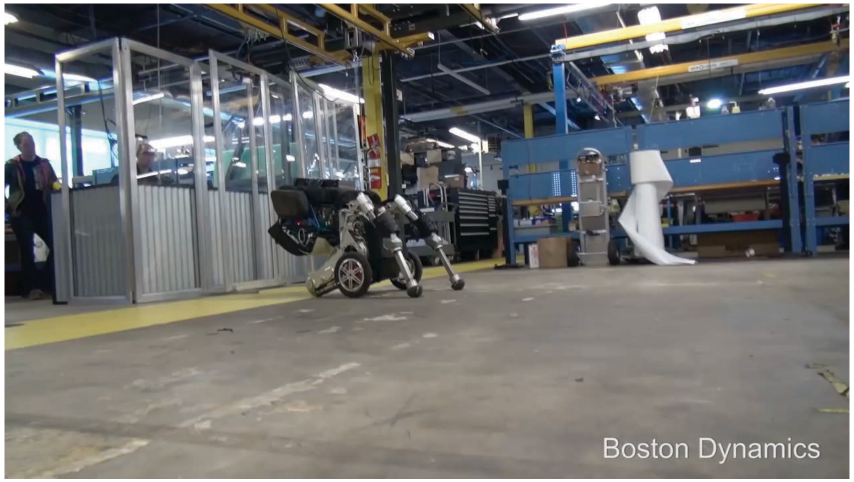




Deep Reinforcement Learning

- **Ex:** AI playing video games, robots, Go game



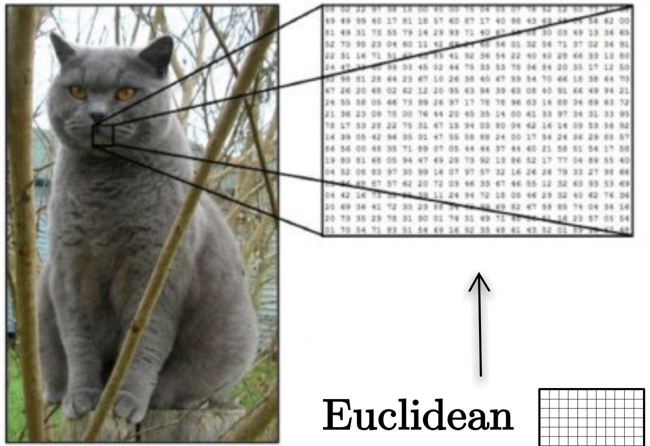


Outline

- History of Deep Learning
- What makes DL great?
- DL architectures
- **Future of DL**
- How to develop AI products?
- How to develop skills/career in AI?
- Conclusion

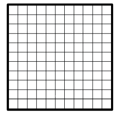
Future of Deep Learning

- Current DL is designed for data lying on **Euclidean grids** ⇒ Everything is mathematically well defined and computationally fast.

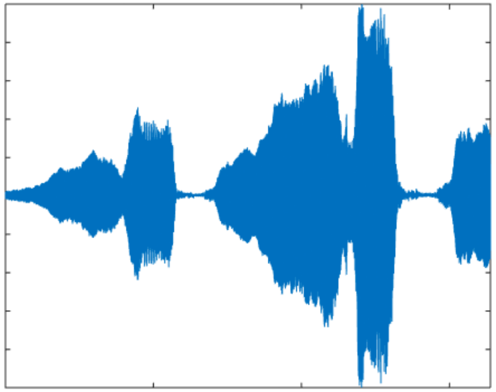


[Karpathy.'15]

Euclidean grid



Images (2D, 3D)
videos (2+1D)



Sound (1D)

- **But not all data lie on Euclidean grids.**

Future of Deep Learning

- **Examples of non-Euclidean data:**

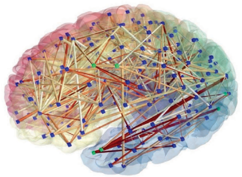
  (1) Social networks (Facebook, Twitter)

  (2) Biological networks (gene, brain connectivity)

  (3) Communication networks (road, wireless, internet)

  (4) Word semantic networks (NLP, translation, Q&A)

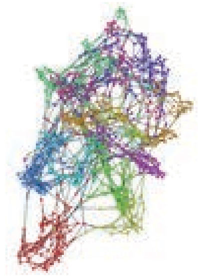
Social networks



Brain structure

Telecommunication networks

=



Graph/network-structured data

- **Generalizing DL to any graph-structured data with *same linear time complexity* as standard DL.**

Future of Deep Learning

Neural Information Processing Systems (NIPS)

Geometric deep learning on graphs and manifolds

M. Bronstein, J. Bruna, A. Szlam, X. Bresson, Y. LeCun

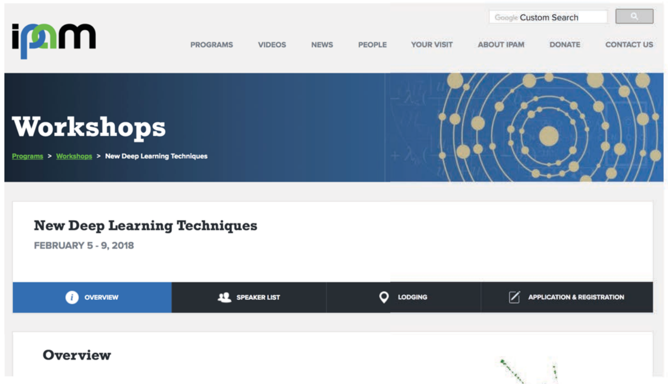
Dec 4, 2017

New Deep Learning Techniques

Feb 5-9, 2018

Organizers

Xavier Bresson, NTU
Michael Bronstein, USI/Intel
Joan Bruna, NYU/Berkeley
Yann LeCun, NYU/Facebook
Stanley Osher, UCLA
Arthur Szlam, Facebook



UCLA

Outline

- History of Deep Learning
- What makes DL great?
- DL architectures
- Future of DL
- **How to develop AI products?**
- How to develop skills/career in AI?
- Conclusion

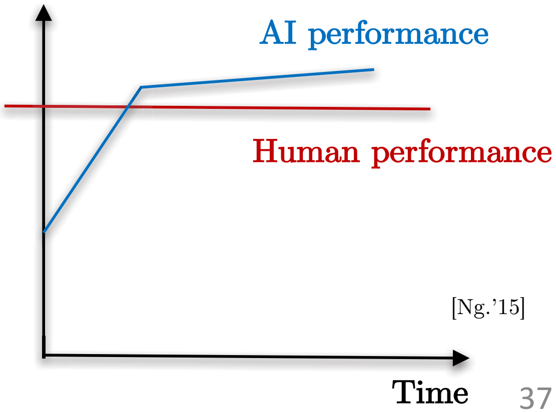
How to develop AI products?

- **What AI can do?** Any task that can be solved by a person with **less than one second of thought**, can be automated with DL [Ng'16].

- **Pipeline of supervised learning products:** Collect data - clean data – label data – test neural network architecture – platform – production.

- Neural network training: **4 numbers are important** – human error, training error, validation error, test error:
  1. If **human=1%, train=8%,** val=9%, test=10%: Collect more data, change architecture.
  2. If human=1%, **train=2%, val=3%, test=10%:** Overfit train-val sets (use regularization), train-test distribution mismatch (redefine train and test sets).

- **Target human error,** because human intelligence is a good **lower bound to intelligence systems** (for vision, hearing, etc, any **perception tasks**).



Outline

- History of Deep Learning
- What makes DL great?
- DL architectures
- Future of DL
- How to develop AI products?
- **How to develop skills/career in AI?**
- Conclusion

How to develop Skills/Career in AI?

- **Online courses:**
  1. ML course of Andrew Ng, Coursera
  2. DL course of Geoffrey Hinton, Coursera
  3. Summer DL course, Montreal 2016, Online
- **Demo AI softwares** (on benchmark datasets):
  1. Google Tensorflow
  2. Facebook PyTorch
- **Experience in the real world:**
  1. Crawl data from Twitter/Facebook
  2. Analyze data statistics
  3. Test DL architectures for a task
- **Kaggle challenge:**
  1. Best practice in ML/DL
  2. Black-box skills
- **Social media news** (follow the stars):
  1. Twitter
  2. Facebook

Outline

- History of Deep Learning
- What makes DL great?
- DL architectures
- Future of DL
- How to develop AI products?
- How to develop skills/career in AI?
- **Conclusion**

Conclusion

- Deep Learning is a **revolutionary** paradigm for **supervised** AI. Super-human performances for object recognition and strategy. Very soon for speech recognition, machine translation, MRI-based disease detection, etc.

- **Not** yet a breakthrough in **non-perceptual** supervised tasks $\Rightarrow$ New techniques are necessary, and have being developed (DL on graphs).

- AI = Math + Data + GPU: **All components are improving very fast ☺.**
  However, performance analysis and understanding are still missing.

- **Are we far away from true AI?** Definition of intelligence needed?
  Probably not, intelligence can be quantified by ability to solve different tasks with a general learning system like the brain (that quantifies intelligence with IQ test).

- **Obstacles to true AI:** Unsupervised learning, generic architecture that transfers to various tasks/simulate the world, reasoning & planning.

Top Influencers

Theory + Algorithms

Geoffrey Hinton
(Toronto, Google)

Yann LeCun
(NewYork, Facebook)

Yoshua Bengio
(Montreal, Microsoft)

Juergen Schmidhuber
(USI, Switzerland)

Industry

Andrew Ng
(Stanford, Baidu)

Jeffrey Dean
(Google)

Demis Hassabis
(DeepMind)

References

- **Talks:**
  - Andrew Ng, Nuts and bolts of building AI applications using Deep Learning, 2016
  - Yann LeCun, Deep Learning and the Future of AI, 2017
  - Rob Fergus, The Deep Learning and Artificial Intelligence Revolution, 2017

- **Articles:**
  - Yoshua Bengio, Ian Goodfellow and Aaron Courville, Deep Learning, 2016, MIT Press
  - Fei-Fei Li, Stanford CS231n: Convolutional Neural Networks for Visual Recognition, 2016.
  - Tran Quoc Hoan, Deep Learning and Business Models, 2013.
  - M. Jordan, Frontiers in Massive Data Analysis, 2001, US NAS
  - Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Deep learning, Nature, 2015
  - Juergen Schmidhuber, Deep learning in neural networks: An overview, 2015xd

Questions?