深度学习技术与应用

Deep Learning: Techniques and Applications

李 戈

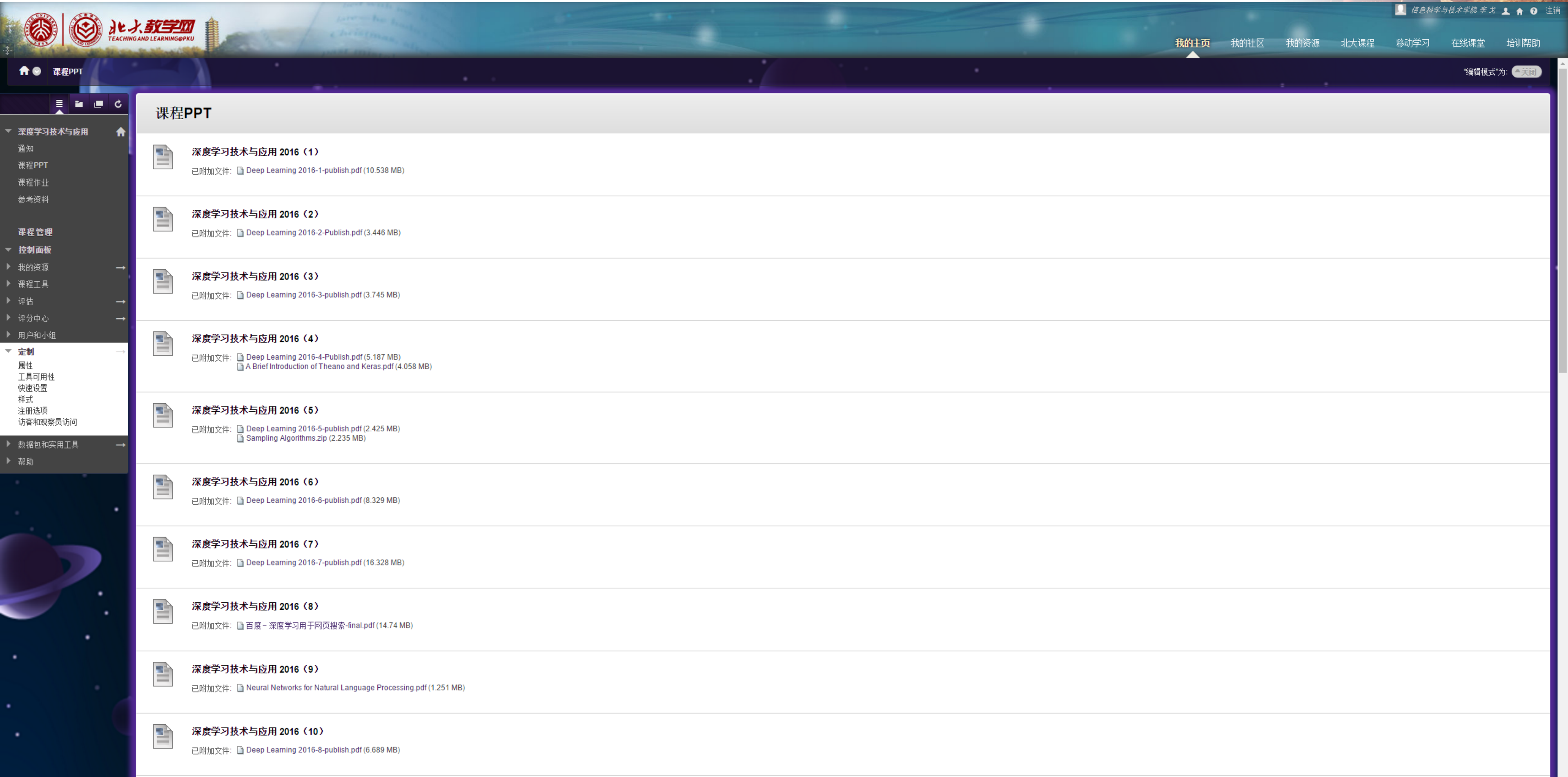
北京大学 信息科学技术学院

lige@pku.edu.cn

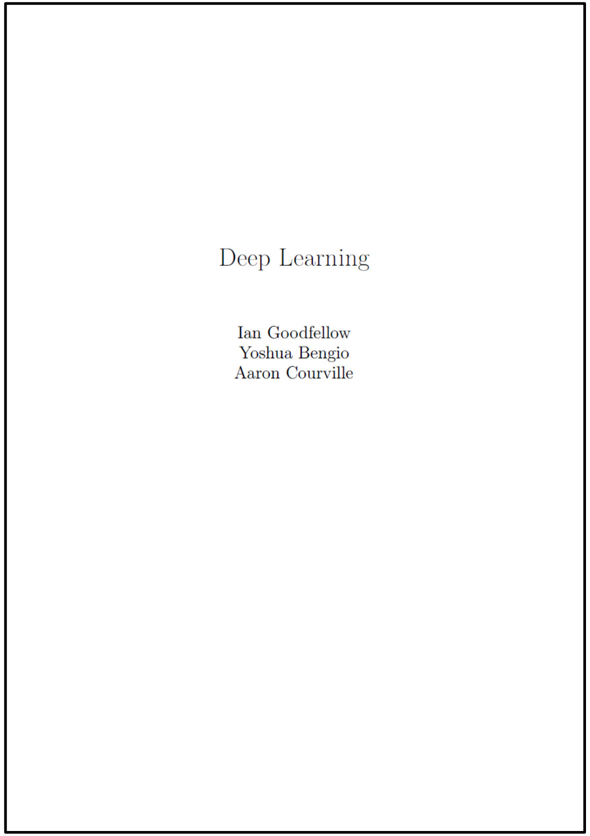
深度学习技术与应用 2017

该二维码7天内(3月2日前)有效，重新进入将更新

course.pku.edu.cn

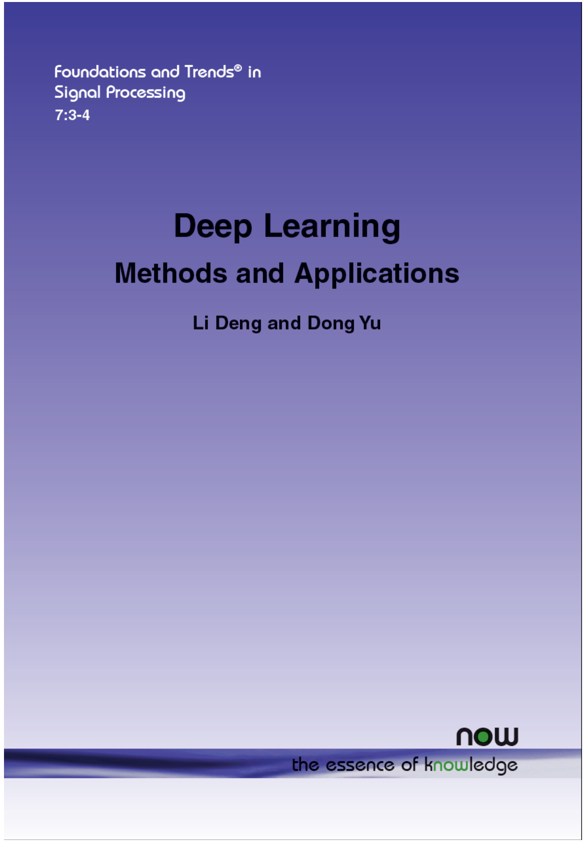


参考资料



Deep Learning

Ian Goodfellow
Yoshua Bengio
Aaron Courville



Foundations and Trends® in
Signal Processing
7:3-4

Deep Learning

Methods and Applications

Li Deng and Dong Yu

now
the essence of knowledge

Neural Networks 61 (2015) 85–117

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

ELSEVIER

Review

Deep learning in neural networks: An overview

Jürgen Schmidhuber

The Swiss AI Lab IDSIA, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, University of Lugano & SUPSI, Galleria 2, 6928 Manno-Lugano, Switzerland

ARTICLE INFO

Article history:
Received 2 May 2014
Received in revised form 12 September 2014
Accepted 14 September 2014
Available online 13 October 2014

Keywords:
Deep learning
Supervised learning
Unsupervised learning
Reinforcement learning
Evolutionary computation

ABSTRACT

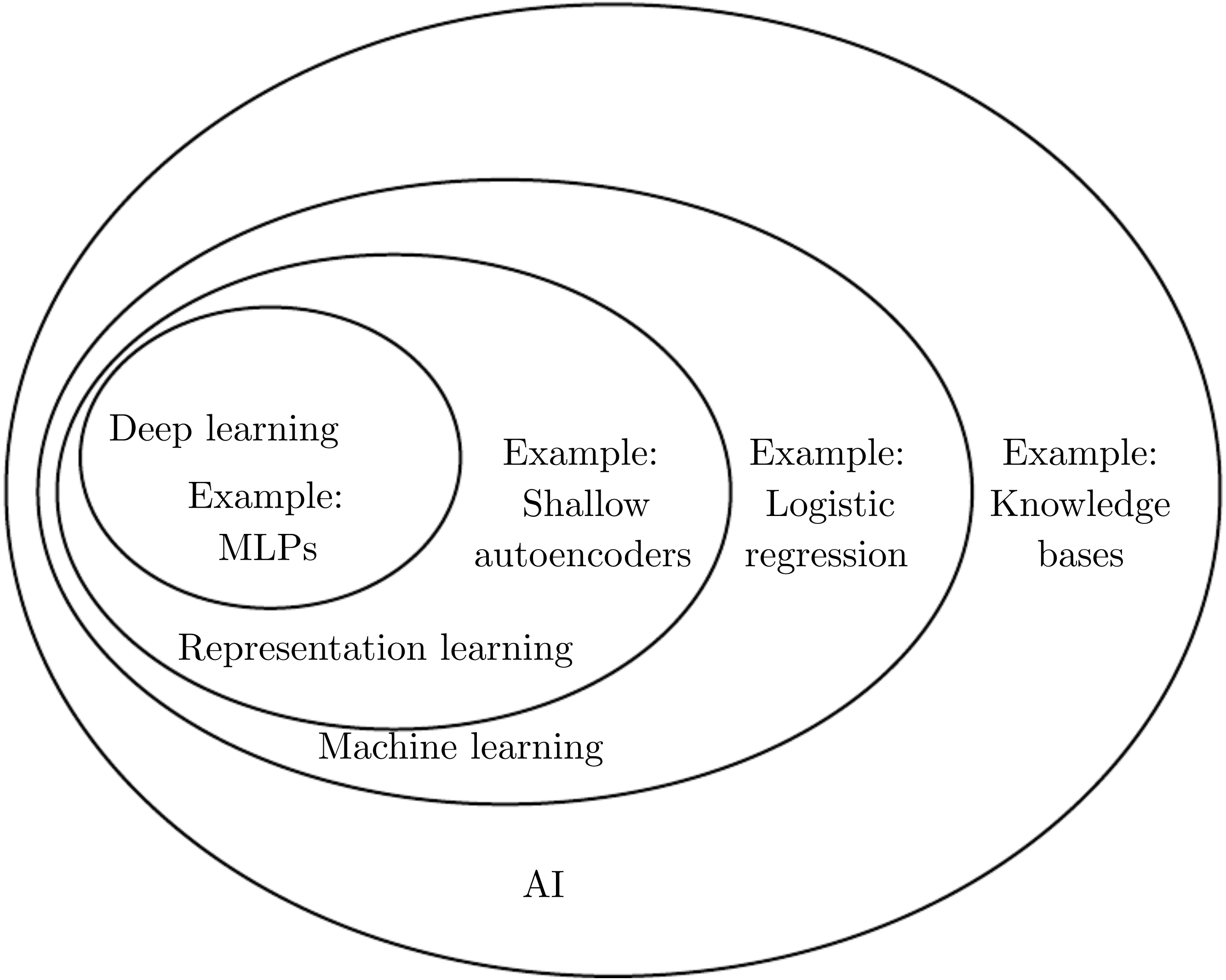
In recent years, deep artificial neural networks (including recurrent ones) have won numerous contests in pattern recognition and machine learning. This historical survey compactly summarizes relevant work, much of it from the previous millennium. Shallow and Deep Learners are distinguished by the depth of their *credit assignment paths*, which are chains of possibly learnable, causal links between actions and effects. I review deep supervised learning (also recapitulating the history of backpropagation), unsupervised learning, reinforcement learning & evolutionary computation, and indirect search for short programs encoding deep and large networks.

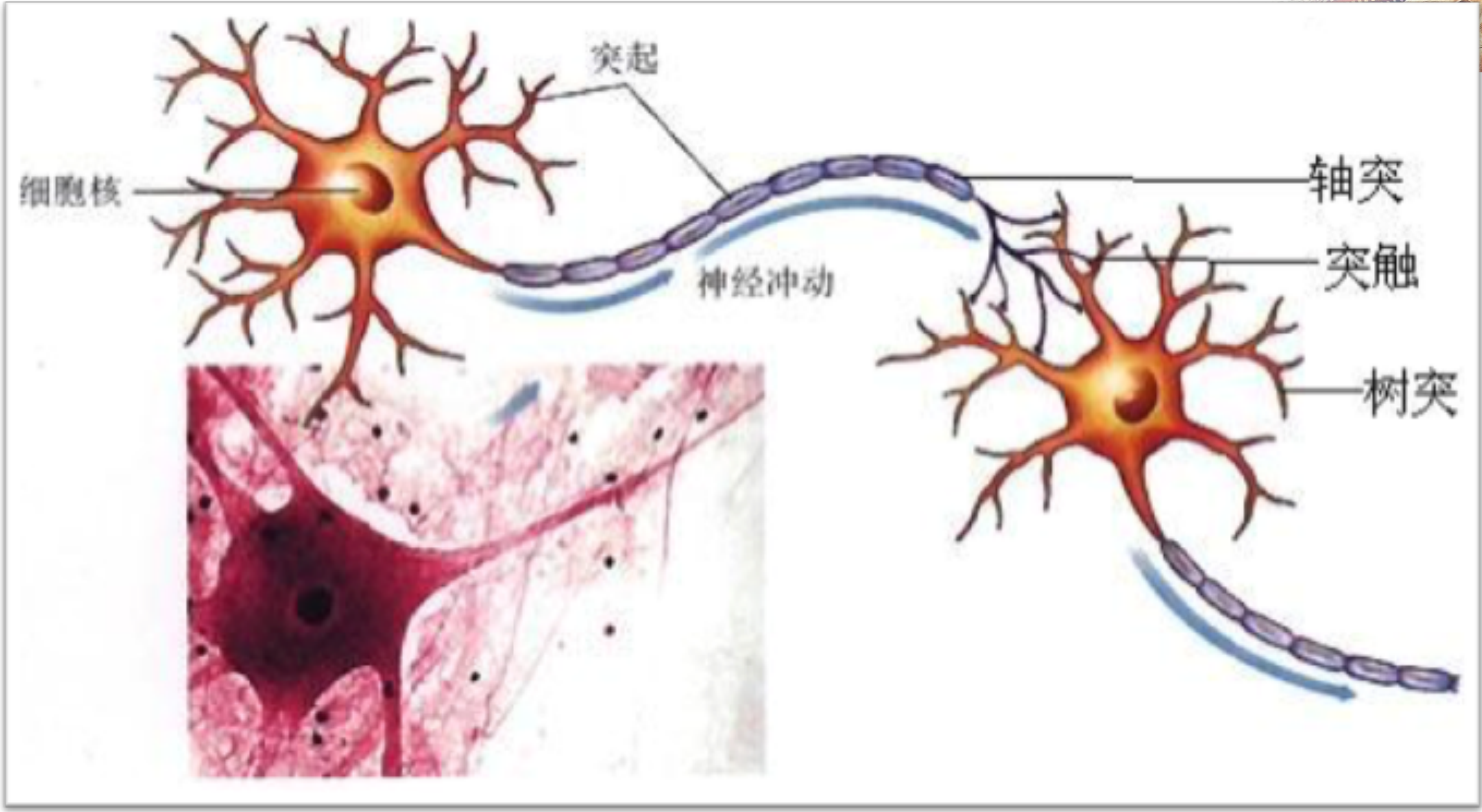
Contents

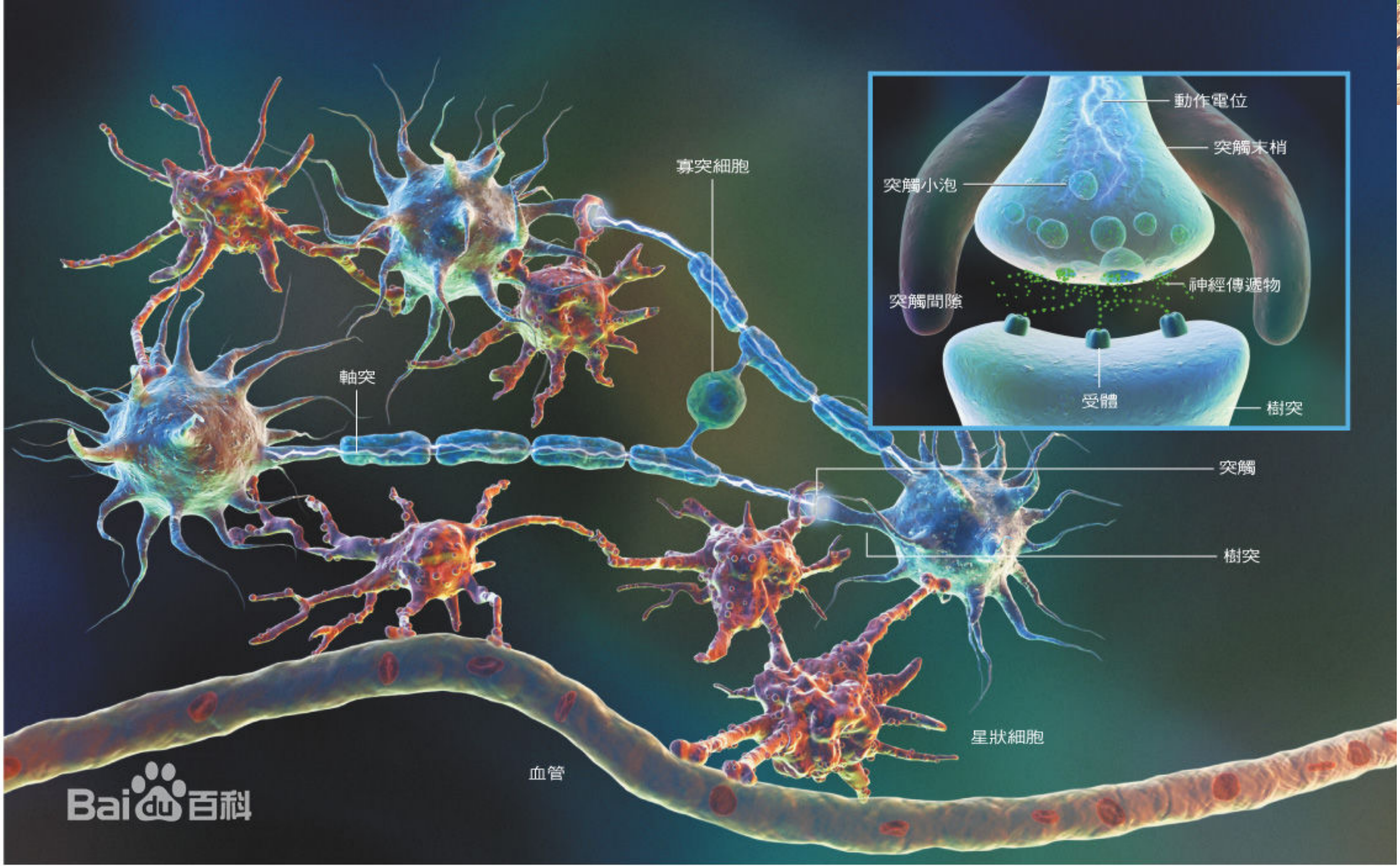
E-mail address: juergen@idsia.ch.

http://dx.doi.org/10.1016/j.neunet.2014.09.003
0893-6080/© 2014 Published by Elsevier Ltd.

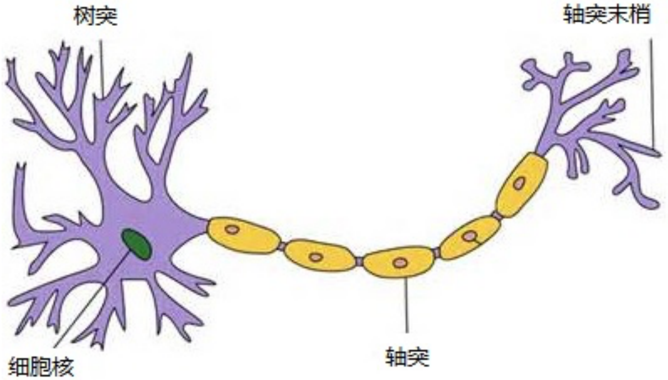
Deep Learning

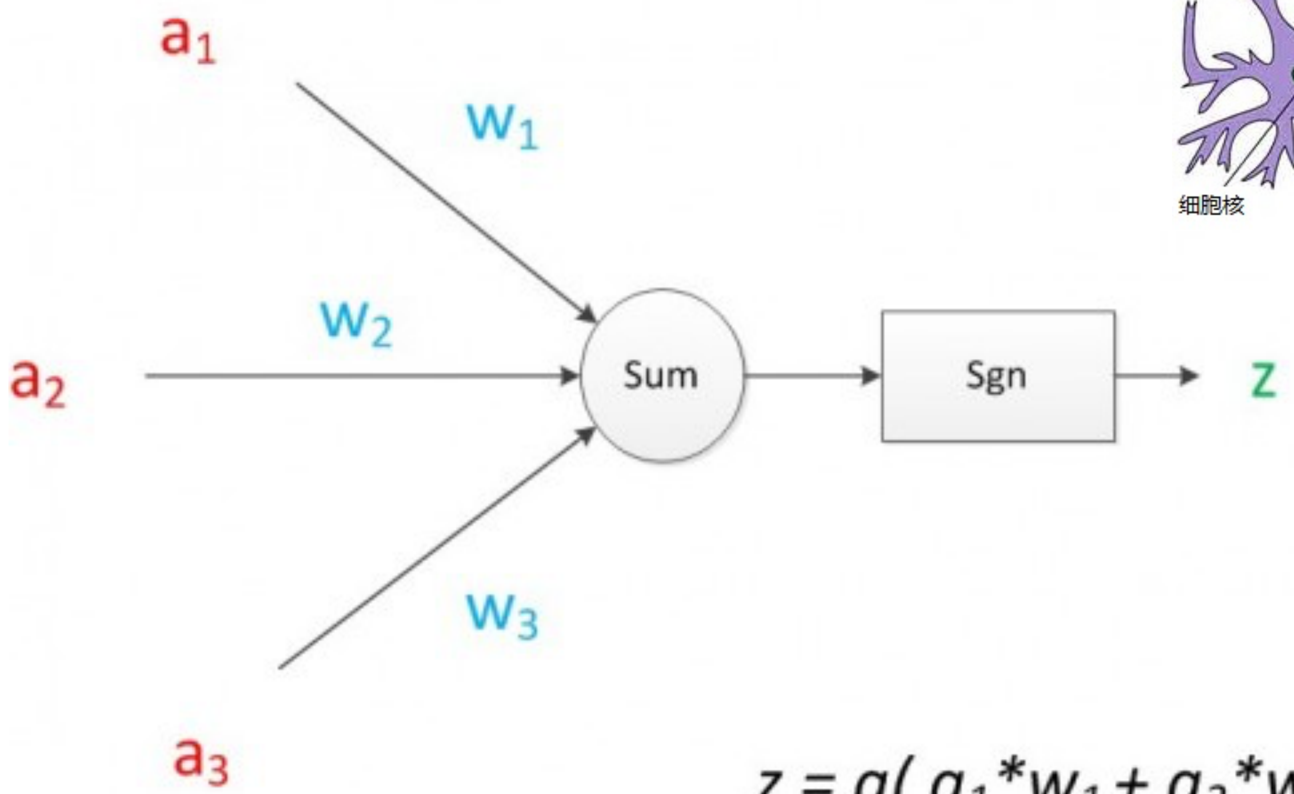



突起
细胞核
轴突
神经冲动
突触
树突


動作電位
突觸末梢
突觸小泡
神經傳遞物
突觸間隙
受體
樹突
寡突細胞
軸突
突觸
樹突
星狀細胞
血管

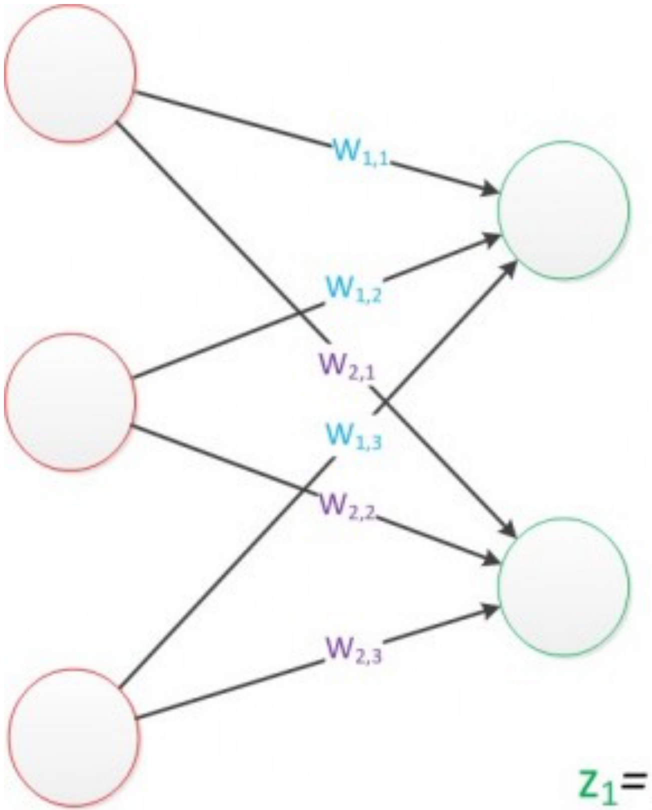
1943

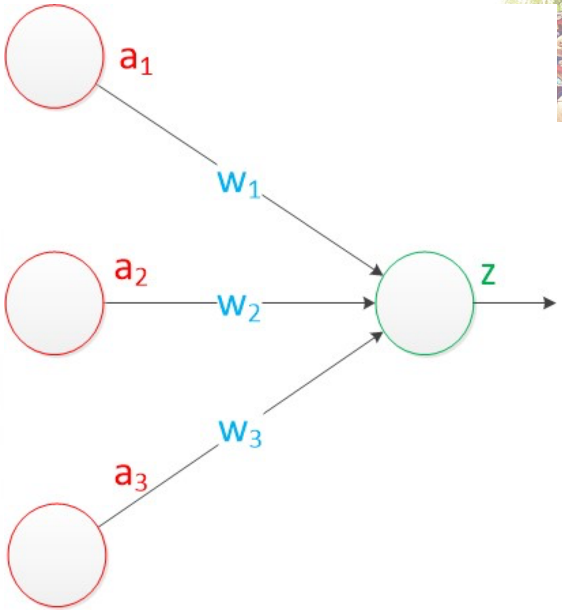




$$z = g(a_1 * w_1 + a_2 * w_2 + a_3 * w_3)$$

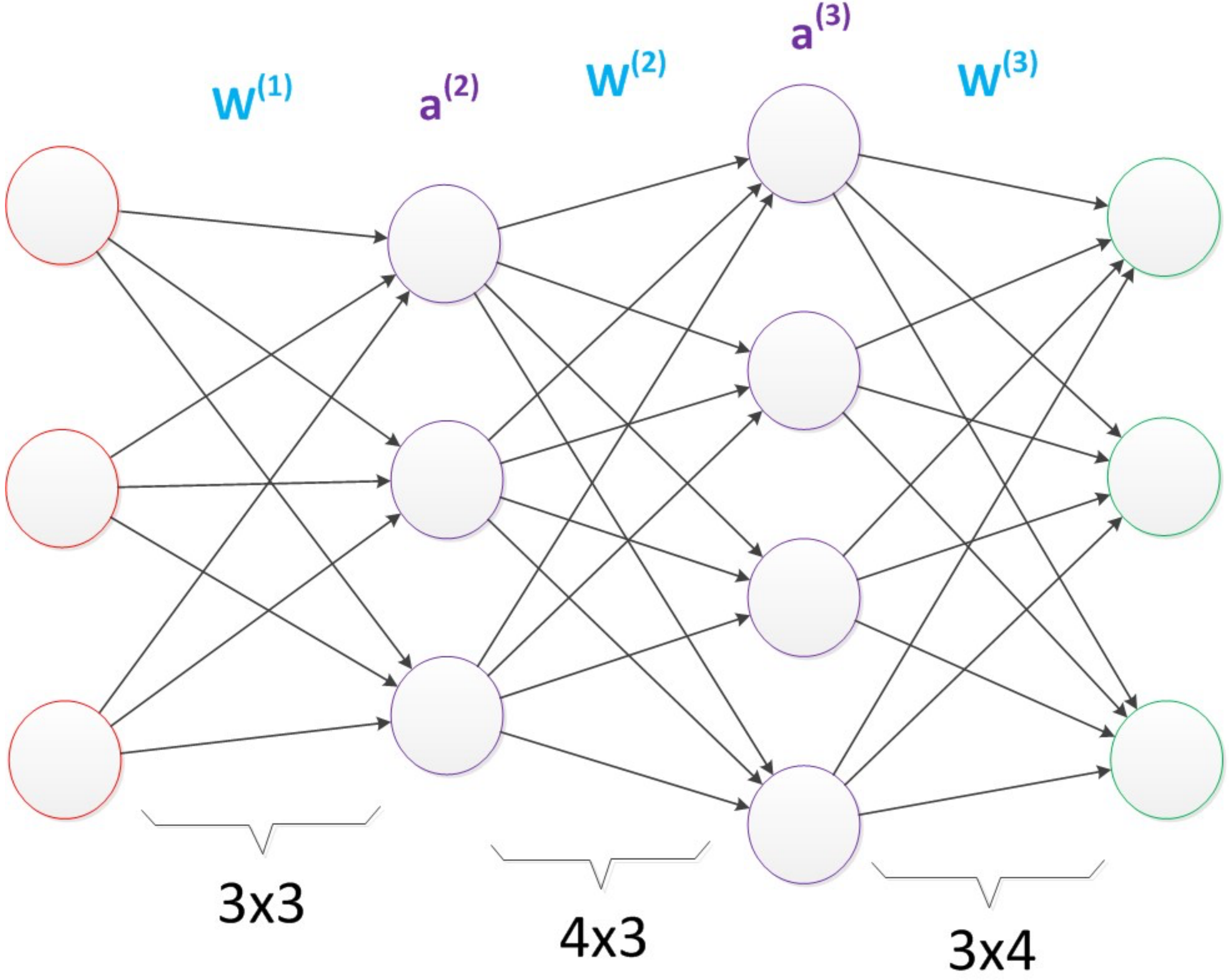
1958





$$z_1 = g(\,a_1 * w_{1,1} + a_2 * w_{1,2} + a_3 * w_{1,3}\,)$$

$$z_2 = g(\,a_1 * w_{2,1} + a_2 * w_{2,2} + a_3 * w_{2,3}\,)$$

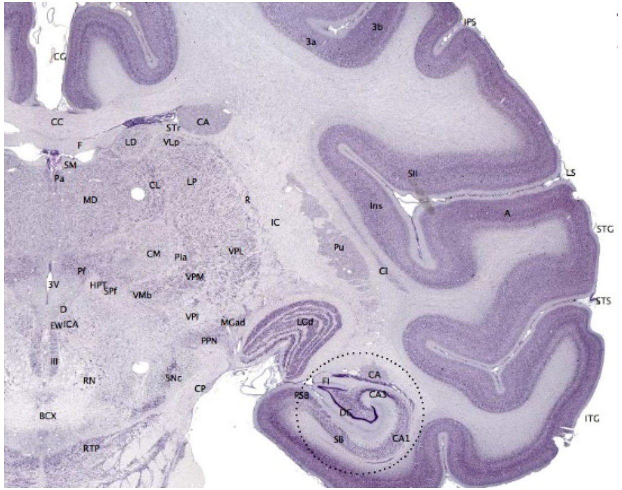


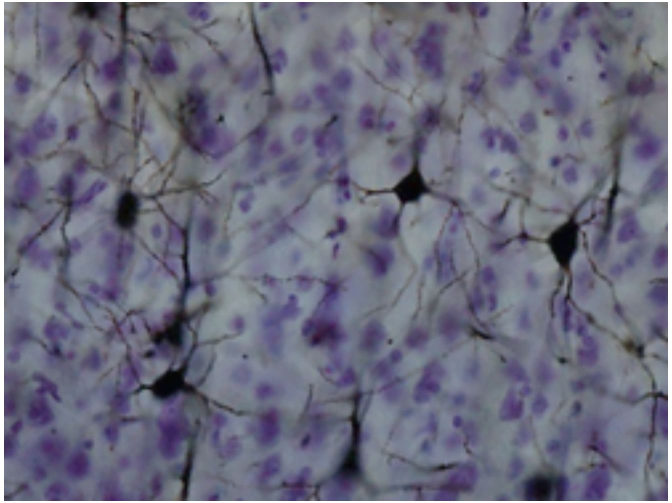
9+12+12=33

Why Deep?

Brain: Cerebral cortex (大脑皮层)

- The human cerebral cortex is 2 to 4 millimetres in thick.
- The different cortical layers each contain a characteristic distribution of neuronal cell types and connections with other cortical and subcortical regions.

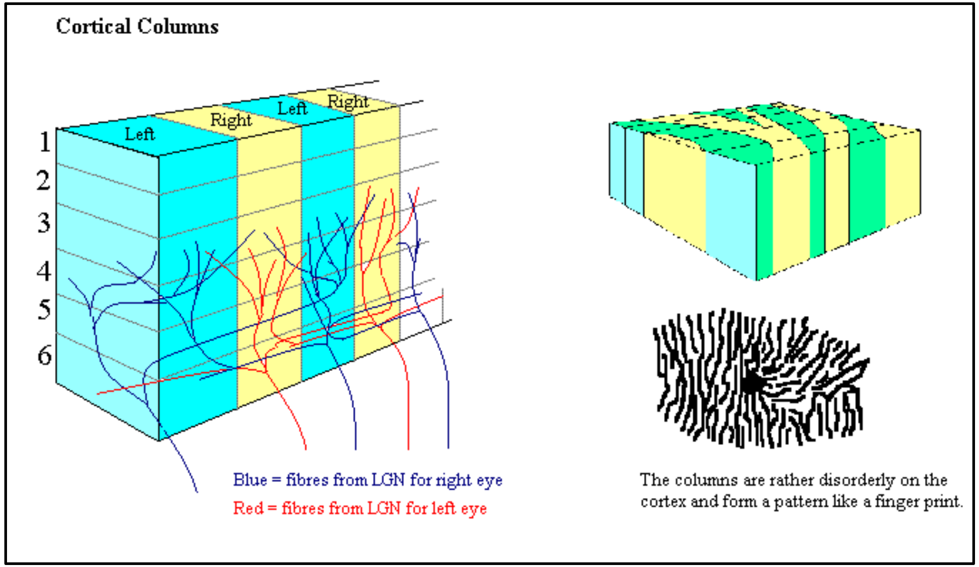




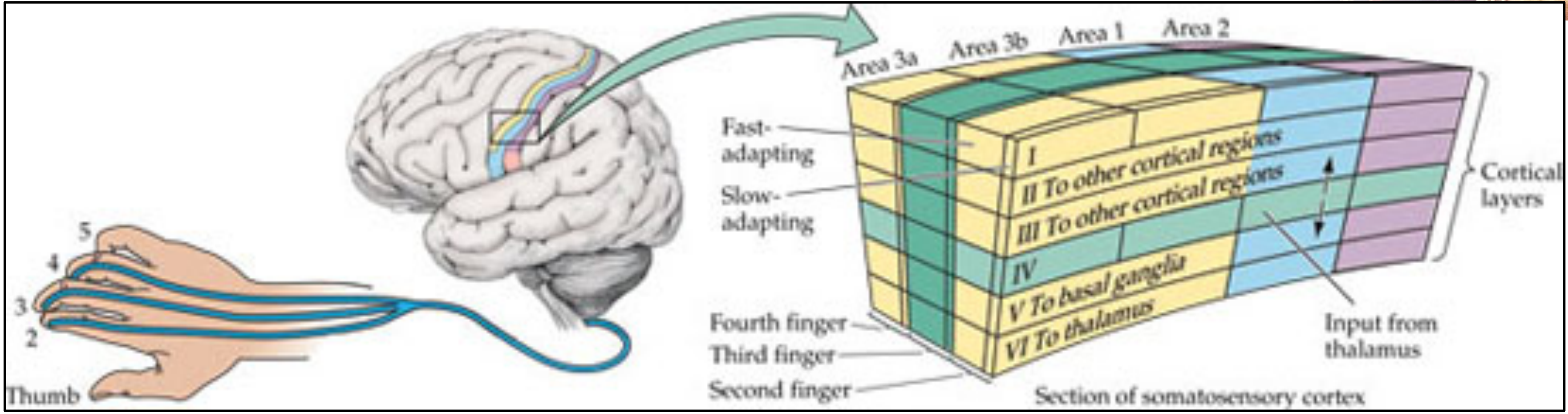
http://en.wikipedia.org/wiki/Cerebral_cortex

Layered structure

- the more ancient part of the cerebral cortex, the hippocampus, has at most three cellular layers.
- The most recent part of the cerebral cortex, the neocortex (also called isocortex), is differentiated into six horizontal layers;
- Neurons in various layers connect vertically to form small microcircuits, called cortical columns.



Layered structure



- The cortex is organized vertically in columns and horizontally in layers.
- The different regions of somatosensory receive their main inputs from different kinds of receptors.
  - Area 3b receives most of its projections from the superficial skin.
  - Area 3a receives input from receptors in the muscle spindles.
- Input from the thalamus arrives at layer IV, where neurons distribute information up and down layers. [Kaas et al., 1979.]

Theories of Brain Development

- Computational deep learning is closely related to a class of theories of brain development (specifically, neocortical development) proposed by cognitive neuroscientists in the early 1990s.[53]
  - An approachable summary of this work is Elman, et al.'s 1996 book "Rethinking Innateness" .
- purely computational deep learning models… appear to be analogous to one way of understanding the neocortex of the brain as a hierarchy of filters…
  - where each layer captures some of the information in the operating environment, and then passes the remainder, as well as modified base signal, to other layers further up the hierarchy.

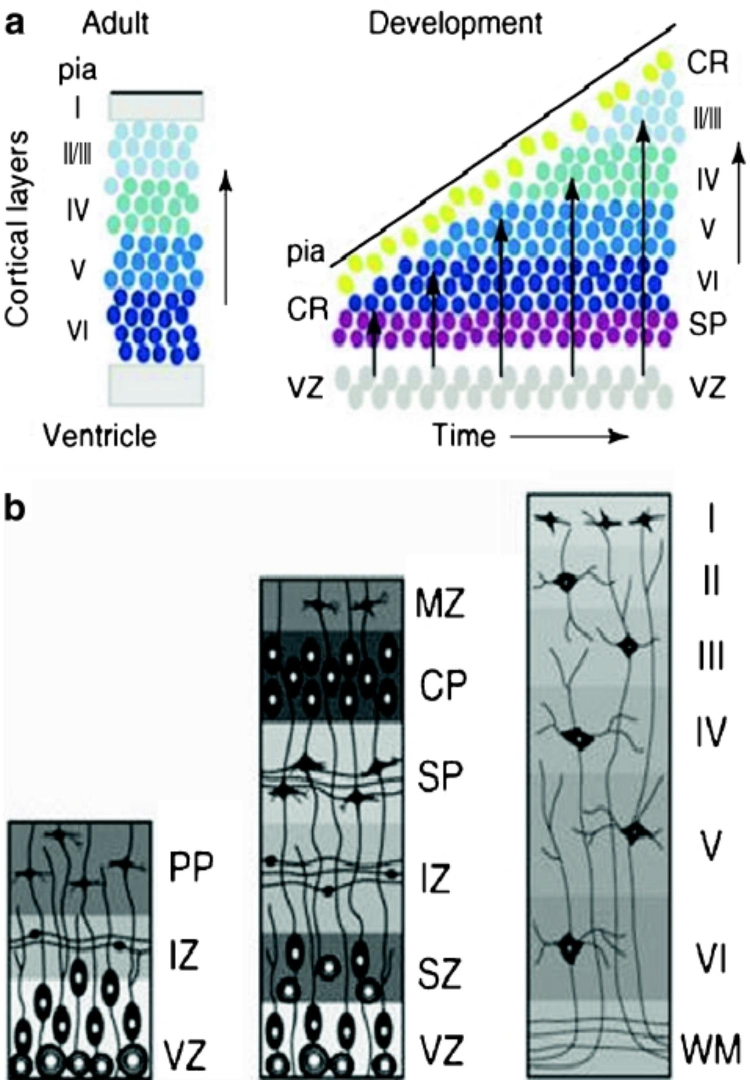
http://en.wikipedia.org/wiki/Deep_learning

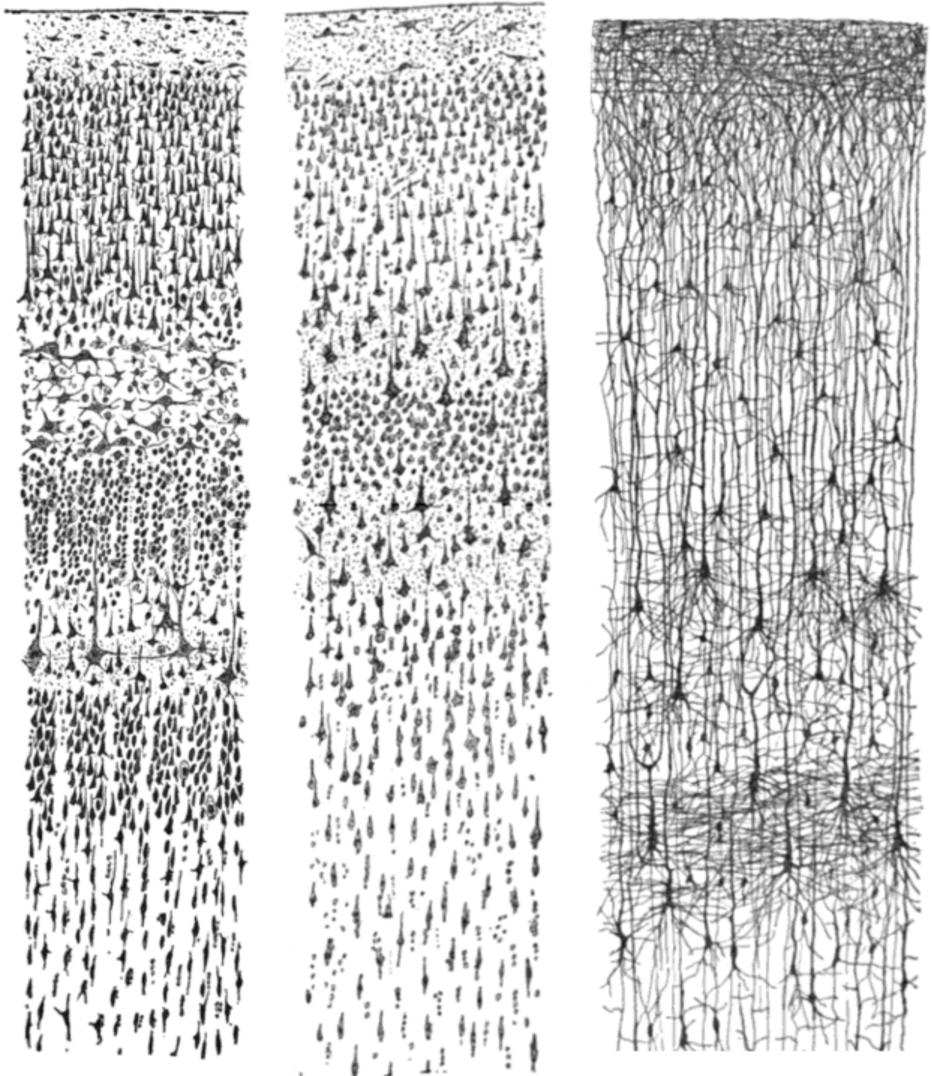
Theories of Brain Development

- One aspect of human development that distinguishes us from our nearest primate neighbors may be changes in the timing of development.
- Among primates, the human brain remains relatively plastic until late in the post-natal period, whereas the brains of our closest relatives are more completely formed by birth.

As described in The New York Times in 1995:
"...the infant's brain seems to organize itself under the influence of waves of so-called trophic-factors ... different regions of the brain become connected sequentially, with one layer of tissue maturing before another and so on until the whole brain is mature."

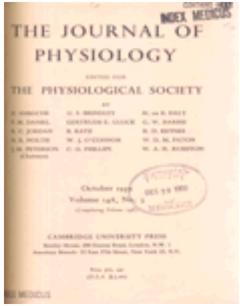
http://en.wikipedia.org/wiki/Deep_learning


a
Adult
Development
pia
I
II/III
IV
V
VI
Cortical layers
Ventricle
CR
II/III
IV
V
VI
SP
VZ
pia
CR
VZ
Time
b
PP
IZ
VZ
MZ
CP
SP
IZ
SZ
VZ
I
II
III
IV
V
VI
WM



Hubel-Wiesel Experiment

- Hubel, David H., and Torsten N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex." *The Journal of physiology* 148, no. 3 (1959): 574-591.


INDEX MEDICUS

THE JOURNAL OF PHYSIOLOGY

THE PHYSIOLOGICAL SOCIETY

CAMBRIDGE UNIVERSITY PRESS

Hubel-Wiesel Experiment

- 简单细胞
  - 光带处于某个空间方位角度时，某些神经元表现活跃；
  - 不同的神经元对不同空间方位的偏好不相同；
  - 神经元对亮光带和暗光带的反应模式不相同；
- 复杂细胞
  - 对于感受野中的边界信息比较敏感；
  - 对于感受野中的运动信息比较敏感；

The Nobel Prize in Physiology or Medicine, 1981

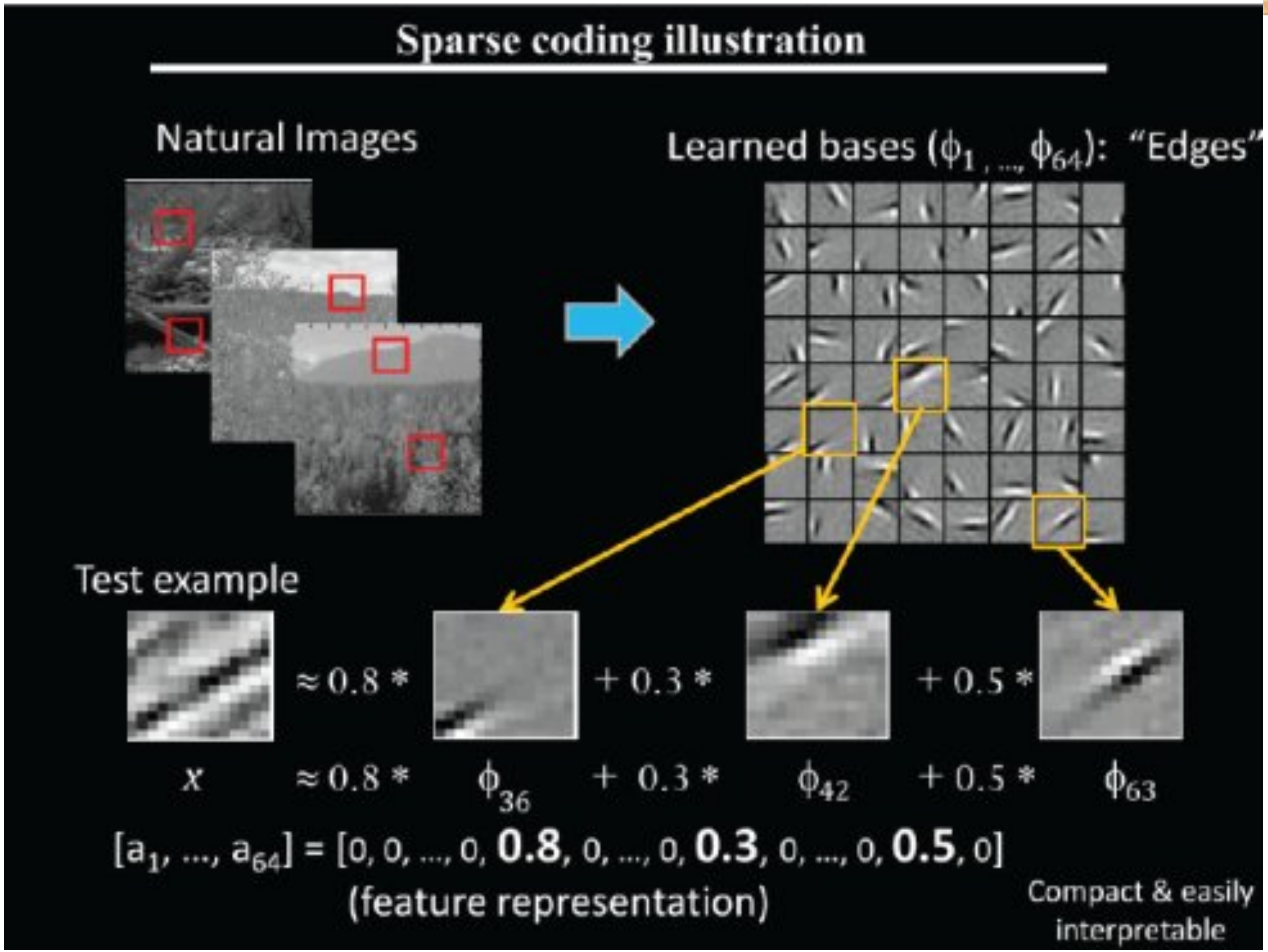
Further Experiment

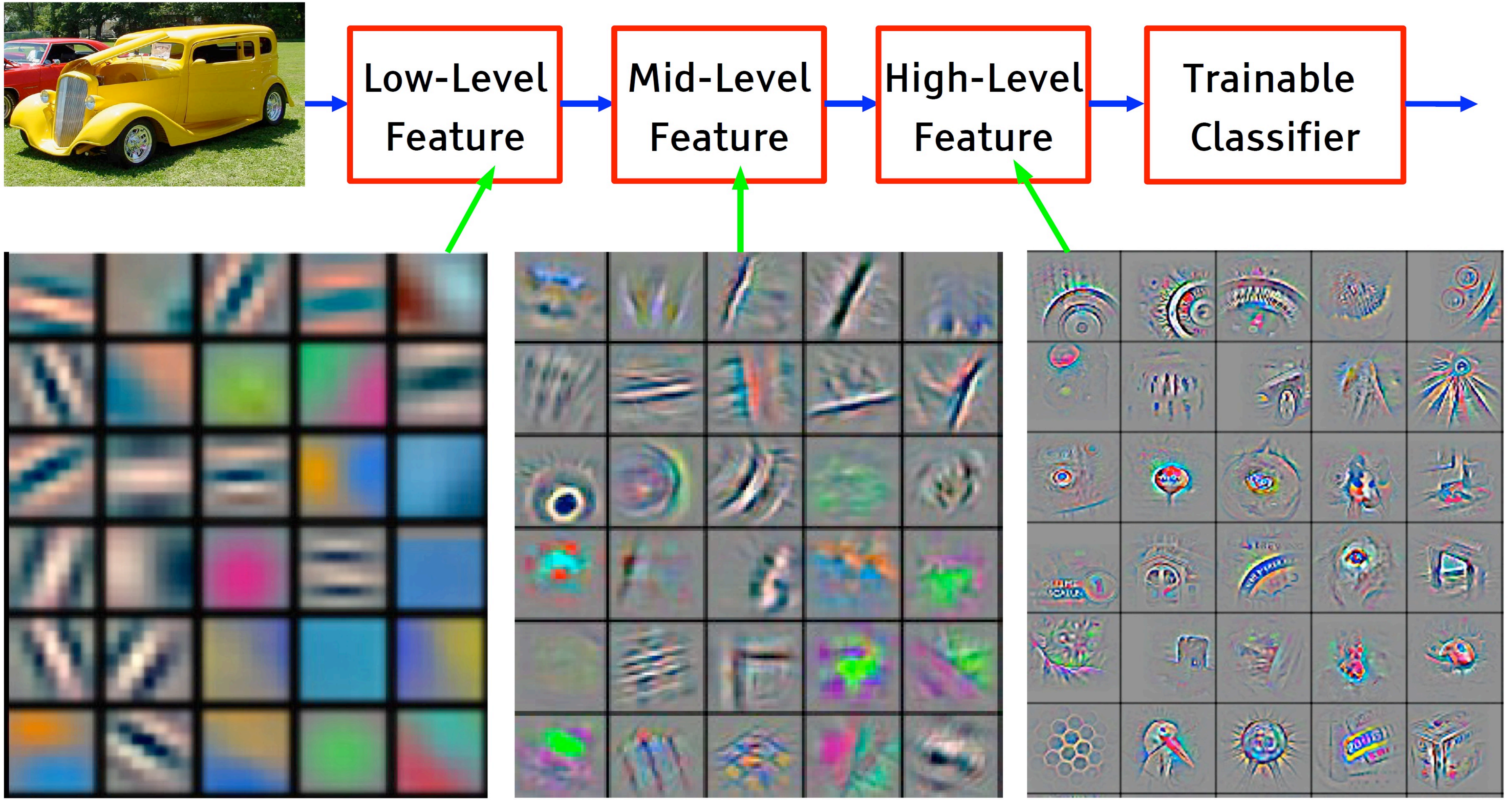
1995

- Bruno Olshausen
- David Field
- Cornell University

从400张碎片中选择尽量少的图片拼接目标图片，发现最终选择的图片，基本是照片上不同物体的边缘。

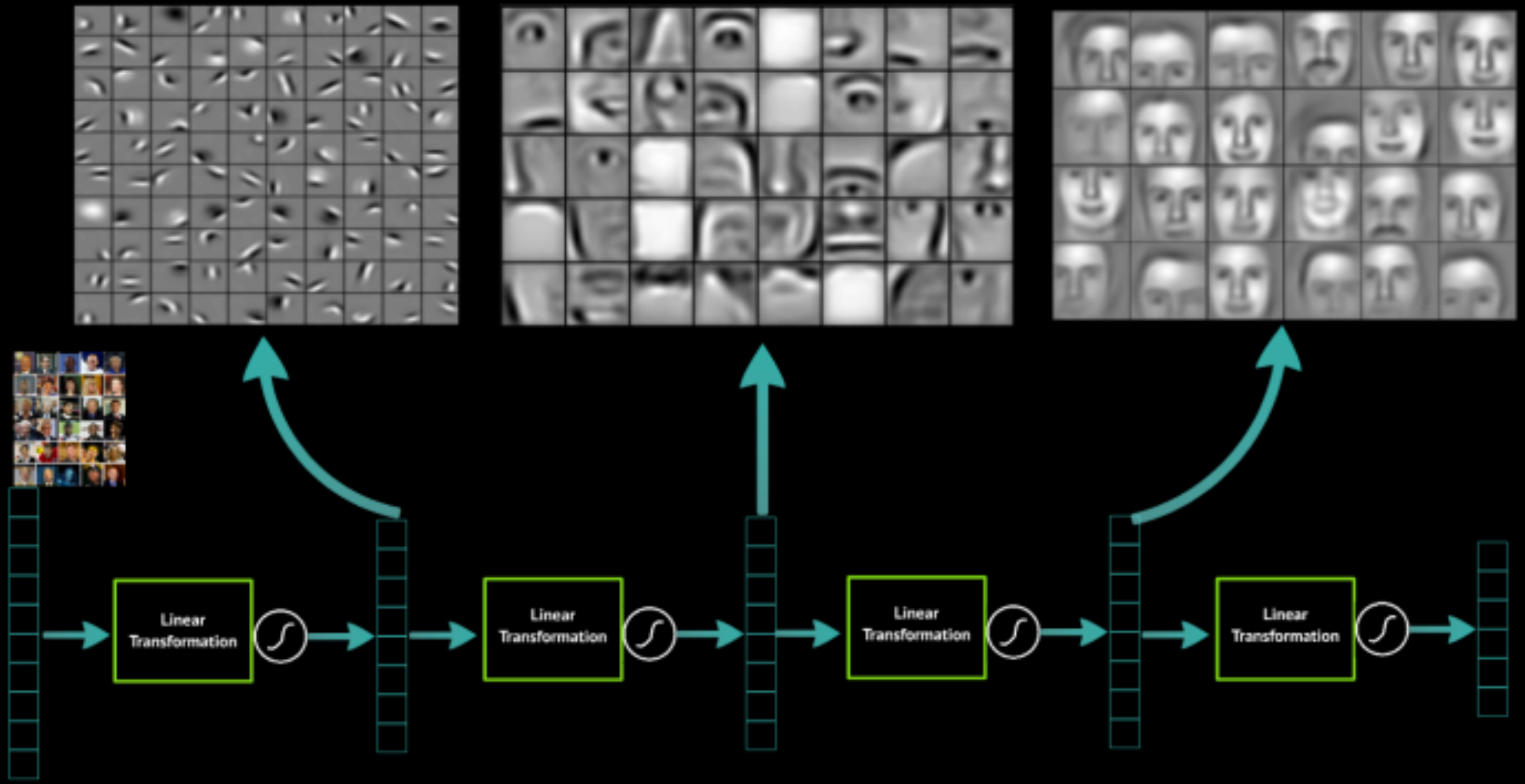
-- 稀疏编码算法





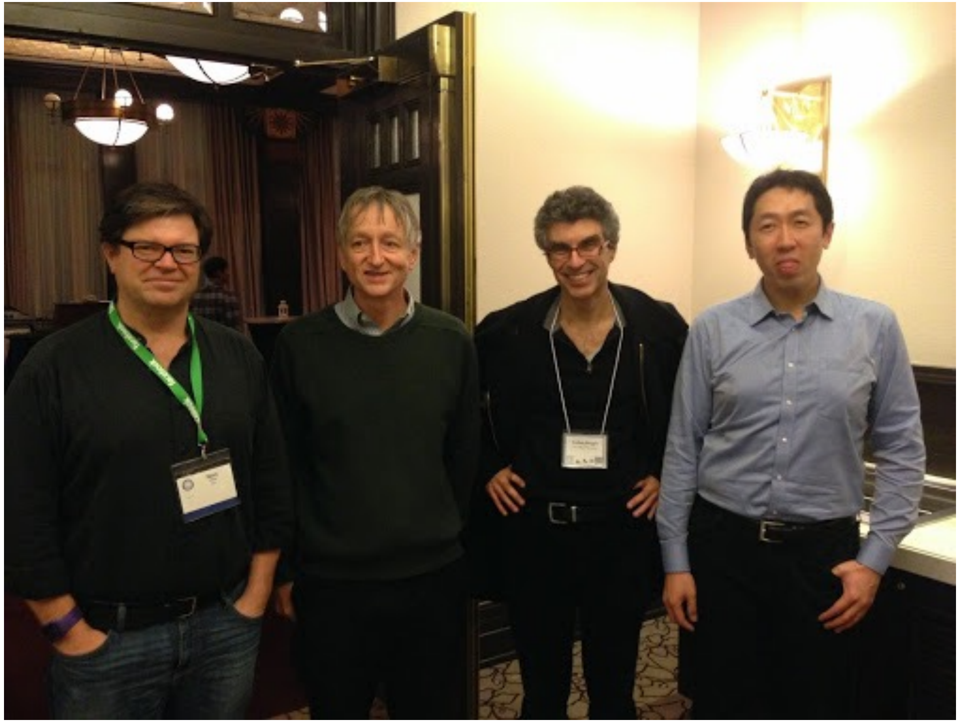
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Deep Learning learns layers of features



- **University of Toronto - Machine Learning Group (Geoff Hinton**, Rich Zemel, Ruslan Salakhutdinov, Brendan Frey, Radford Neal)
- **Université de Montréal - Lisa Lab (Yoshua Bengio**, Pascal Vincent, Aaron Courville, Roland Memisevic)
- **New York University – Yann Lecun**'s and Rob Fergus' group
- **Stanford University – Andrew Ng**'s group
- University of Oxford – Deep learning group, Nando de Freitas and Phil Blunsom
- **Google Research** – Jeff Dean, Samy Bengio, Jason Weston, Marc'Aurelio Ranzato, Dumitru Erhan, Quoc Le et al
- **Microsoft Research** – Li Deng et al
- SUPSI – IDSIA (Jurgen Schmidhuber's group)
- UC Berkeley – Bruno Olshausen's group
- University of Washington – Pedro Domingos' group
- IDIAP Research Institute - Ronan Collobert's group
- University of California Merced – Miguel A. Carreira-Perpinan's group
- University of Helsinki - Aapo Hyvärinen's Neuroinformatics group
- Université de Sherbrooke – Hugo Larochelle's group
- University of Guelph – Graham Taylor's group
- University of Michigan – Honglak Lee's group
- Technical University of Berlin – Klaus-Robert Muller's group
- **Baidu – Kai Yu's group**
- Aalto University - Juha Karhunen and Tapani Raiko group
- U. Amsterdam – Max Welling's group
- U. California Irvine – Pierre Baldi's group
- Ghent University – Benjamin Shrauwen's group
- University of Tennessee – Itamar Arel's group
- **IBM Research – Brian Kingsbury et al**
- University of Bonn – Sven Behnke's group
- Gatsby Unit @ University College London – Maneesh Sahani, Yee-Whye Teh, Peter Dayan
- Computational Cognitive Neuroscience Lab @ University of Colorado Boulder

University of Toronto - **Machine Learning Group** (Geoffrey Hinton, Rich Zemel, Ruslan Salakhutdinov, Brendan Frey, Radford Neal)
Université de Montréal – **MILA Lab** (Yoshua Bengio, Pascal Vincent, Aaron Courville, Roland Memisevic)
New York University – **Yann Lecun**, **Rob Fergus**, David Sontag and Kyunghyun Cho
Stanford University – **Andrew Ng**, **Christopher Manning**'s, **Fei-fei Li**'s group
University of Oxford – **Deep learning group**, **Nando de Freitas** and **Phil Blunsom**, Andrew Zisserman
Google Research – Jeff Dean, Geoffrey Hinton, Samy Bengio, Ilya Sutskever, Ian Goodfellow, Oriol Vinyals, Dumitru Erhan, Quoc Le et al
Google DeepMind - Alex Graves, Karol Gregor, Koray Kavukcuoglu, Andriy Mnih, Guillaume Desjardins, Xavier Glorot, Razvan Pascanu, Volodymyr Mnih et al
Facebook AI Research(FAIR) - Yann Lecun, Rob Fergus, Jason Weston, Antoine Bordes, Soumit Chintala, Leon Bouttou, Ronan Collobert, Yann Dauphin et al.
Twitter's Deep Learning Group – Hugo Larochelle, Ryan Adams, Clement Farabet et al
Microsoft Research – **Li Deng** et al
SUPSI – **IDSIA** (**Jurgen Schmidhuber**'s group)
UC Berkeley – **Bruno Olshausen**'s group, **Trevor Darrell**'s group, **Pieter Abbeel**
UCLA – **Alan Yuille**
University of Washington – **Pedro Domingos**' group
IDIAP Research Institute - **Ronan Collobert**'s group
University of California Merced – **Miguel A. Carreira-Perpinan**'s group
University of Helsinki - **Aapo Hyvärinen**'s Neuroinformatics group
Université de Sherbrooke – **Hugo Larochelle**'s group
University of Guelph – **Graham Taylor**'s group
University of Michigan – **Honglak Lee**'s group
Technical University of Berlin – **Klaus-Robert Muller**'s group
Baidu – **Kai Yu**'s and Andrew Ng's group
Aalto University - **Juha Karhunen and Tapani Raiko** group
U. Amsterdam – **Max Welling**'s group
CMU – **Chris Dyer**
U. California Irvine – **Pierre Baldi**'s group
Ghent University – **Benjamin Shrauwen**'s group
University of Tennessee – **Itamar Arel**'s group
IBM Research – **Brian Kingsbury** et al
University of Bonn – **Sven Behnke's** group
Gatsby Unit @ University College London – Maneesh Sahani, Peter Dayan
Computational Cognitive Neuroscience Lab @ University of Colorado Boulder
DBsystem group @ National University of Singapore



Software libraries

TensorFlow — Google's open source machine learning library in C++ and Python with APIs for both. It provides parallelization with CPUs and GPUs.

Theano — An open source machine learning library for Python supported by the University of Montreal and Yoshua Bengio's team.

Torch — An open source software library for machine learning based on the Lua programming language and used by Facebook.

Caffe - Caffe is a deep learning framework made with expression, speed, and modularity in mind. It is developed by the Berkeley Vision and Learning Center (BVLC) and by community contributors.

Deeplearning4j — An open-source deep-learning library written for Java/C++ with LSTMs and convolutional networks. It provides parallelization with Spark on CPUs and GPUs.

Gensim — A toolkit for natural language processing implemented in the Python programming language.

Keras — An open-source deep learning framework for the Python programming language.

Microsoft CNTK (Computational Network Toolkit) — Microsoft's open-source deep-learning toolkit for Windows and Linux. It provides parallelization with CPUs and GPUs across multiple servers.

MXNet — An open source deep learning framework that allows you to define, train, and deploy deep neural networks.

OpenNN — An open source C++ library which implements deep neural networks and provides parallelization with CPUs.

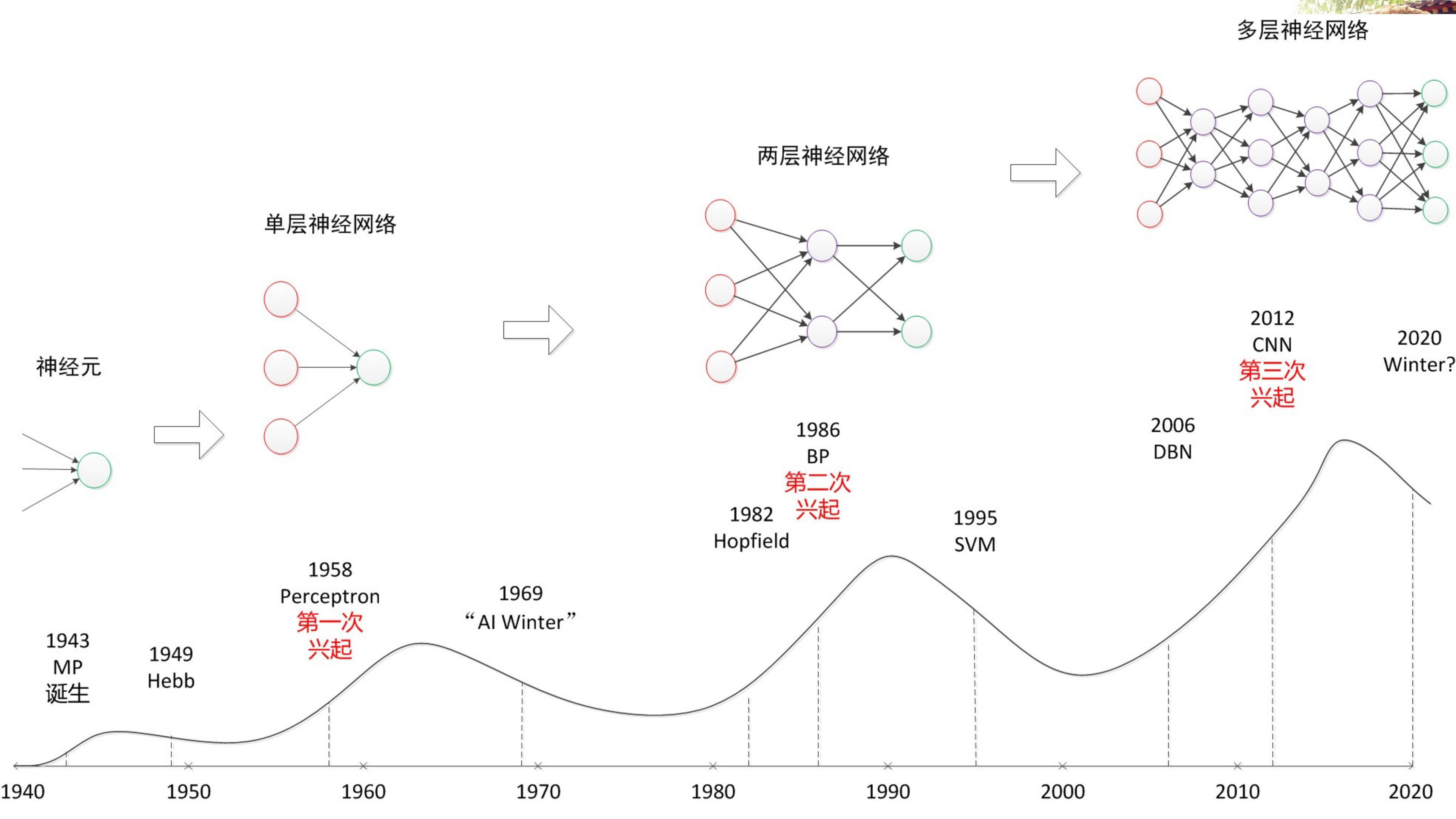
PaddlePaddle — An open source C++ /CUDA library with Python API for scalable deep learning platform with CPUs and GPUs, originally developed by Baidu.

DIANNE - A modular open-source deep learning framework in Java / OSGi developed at Ghent University, Belgium. It provides parallelization with CPUs and GPUs across multiple servers.

Historical Context of Deep Learning

There's an interesting history about people's changes in their attitudes toward the deep architectures and the shallow architectures.

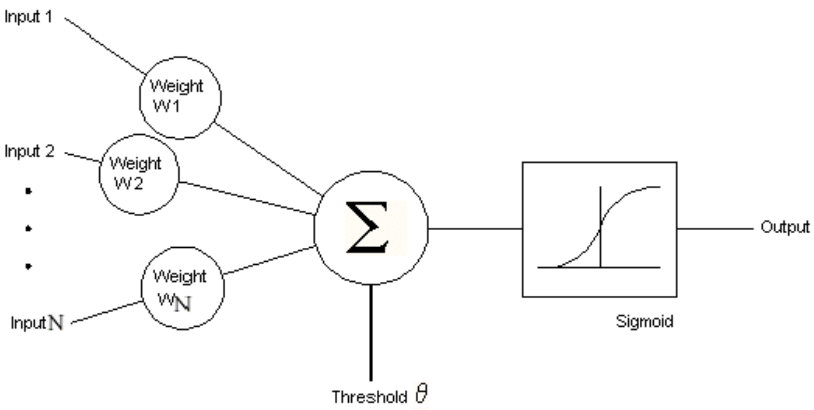
2012 - A survey on deep learning - one small step toward AI

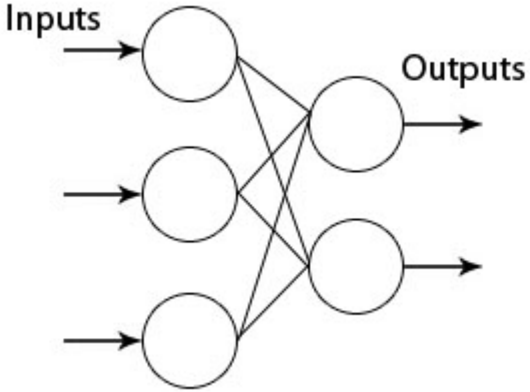


神经元
单层神经网络
两层神经网络
多层神经网络
1943
MP
诞生
1949
Hebb
1958
Perceptron
第一次
兴起
1969
“AI Winter”
1982
Hopfield
1986
BP
第二次
兴起
1995
SVM
2006
DBN
2012
CNN
第三次
兴起
2020
Winter?
1940
1950
1960
1970
1980
1990
2000
2010
2020

Perceptron

- In around 1960, the 1st generation of neural network was born(by Frank Rosenblatt).
- It's capability of classifying some basic shapes like triangles and squares let people see the potential that a real intelligent machine which can sense, learn, remember and recognize like human-beings can be invented with this trend.





Perceptron

- BUT, its fundamental limitations soon broke people's dreams.
- Criticizing from Marvin Minsky, 1969
  - One of the apparent reasons is that the feature layer of this Perceptron is fixed and crafted by human beings, which is absolutely against the definition of a real “intelligent" machine.
  - Another reason is its single-layer structure limits the functions it can learn, e.g, an exclusive-or function is out of its learning ability,

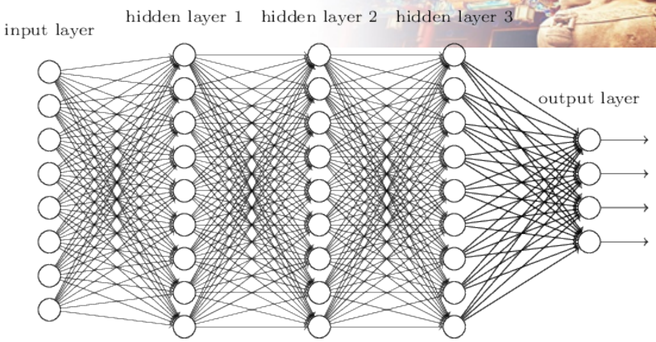
2nd-generation neural network

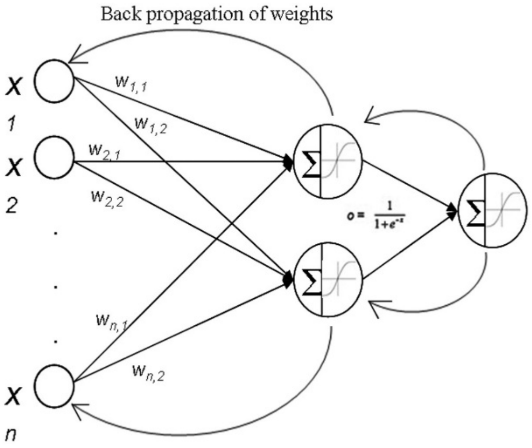
- In around 1985, based on the Perceptrons, Geoffrey Hinton replaced the original single fixed feature layer with several hidden layers, creating the 2nd-generation neural network.
  - via Back-propagation algorithm (proposed in 1969, practicable in 1974)
- In 1989, Yann LeCun et al. built a deep neural network with the purpose of recognizing handwritten ZIP codes on mail.
  - Despite the success of applying the algorithm, the time to train the network on this dataset was approximately 3 days.

Learning internal representations by error propagation
DE Rumelhart, GE Hinton, RJ Williams - 1985 - DTIC Document
... PERSONAL AUTHOR(S) David E. Rumelhart, Geoffrey E. **Hinton**, and Ronald J. Williams 13a ... Continue on reverse if necessary and identify by block number) FIELD GROUP SUB-GROUP -learning; networks; perceptrons; adaptive systems; learning machines; **back propagation** ...

Deep Neural Networks

- **Feed-forward neural networks with many hidden layers**
  - **The weights are initialized of with random weights.**
  - **Fine-tuning by back-propagation algorithm.**
- BP did not work well in practice
  - It often gets **trapped in poor local optima** when the batch-mode or even stochastic gradient descent BP algorithm is used.
  - The **severity increases significantly** as the depth of the networks increases.





2nd-generation neural network

- it still has four main disadvantages as the followings:
  - Lacks the ability to train the unlabeled data while in practice most data is unlabeled;
  - The correcting signal will be weakened when it passes back via multiple layers;
  - Learning is too slow across multiple hidden layers;
  - It can get stuck in poor local optima.
  - Depends too much on human's using experience and skills.

SVM slowed down the developments of NN

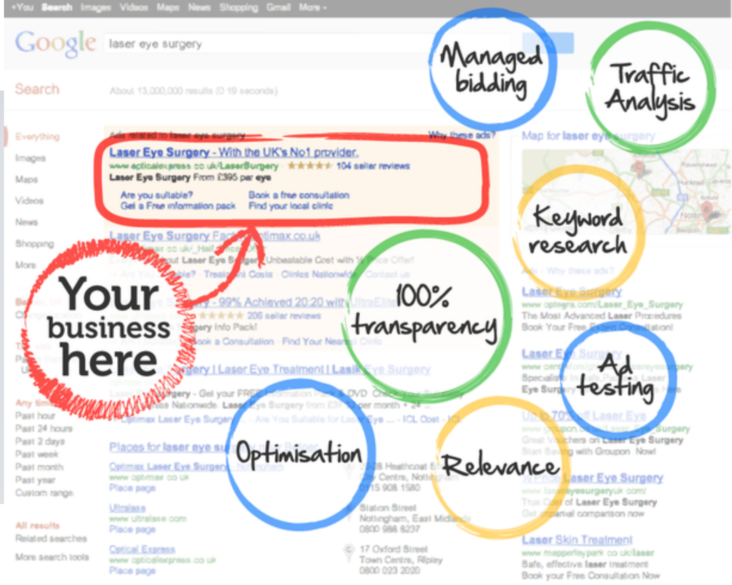
- When people were trying to make improvements to Hinton's neural networks with respect to those advantages
  - Trying to increase the training data set and estimating the initial weight values…
- 1993-1995 Vladimir N. Vapnik, et al.

  made improvements on the original perceptron.
- They creating a new family called

  Support Vector Machines(SVM).

SVM


Google
AdWords


Managed bidding
Traffic Analysis
Keyword research
Your business here
100% transparency
Ad testing
Optimisation
Relevance


百度学习中心
消费情况概括
全账户整体表现
重要事件即时提醒
重点关注对象表现
信息传递窗口


Bai du 推广
专业推广效果升级
百度搜索推广专业版全面上线
首页
功能优势
快速上手
常见问题
全国培训
了解专业版全新功能
学习专业版操作方法
解答专业版常见问题
登录百度推广管理系统
专业推广 轻松掌握
www.itxinwen.com

SVM——Good or Bad?

- SVM makes learning fast and easy, due to its simple structures.
- Appropriate for data with simple structures, e.g, with a small number of features or the data which doesn't contain hierarchical structures.
- But, for the data which itself contains complicated features, SVM tends to perform worse because of its simple structure.
- One way to solve this problem is **to add a prior knowledge** to the SVM model in order to obtain a better feature layer.
- But, it's hard to find a general set of prior knowledge;

SVM——takes us away from the road to a real intelligent machine

- Despite the fact that SVM can work really well in solving many AI problems, it is not a good trend to AI due to its fatal deficiency, shallow architecture.
  - SVM is still a kind of Perceptron where **the features are directly obtained but not learnt from the data itself.**
- With the purpose of finding an architecture that meets the requirements above, some researchers started to look back to the multi-layer neural network, trying to exploit its advantages related to deep and overcome the limitations...

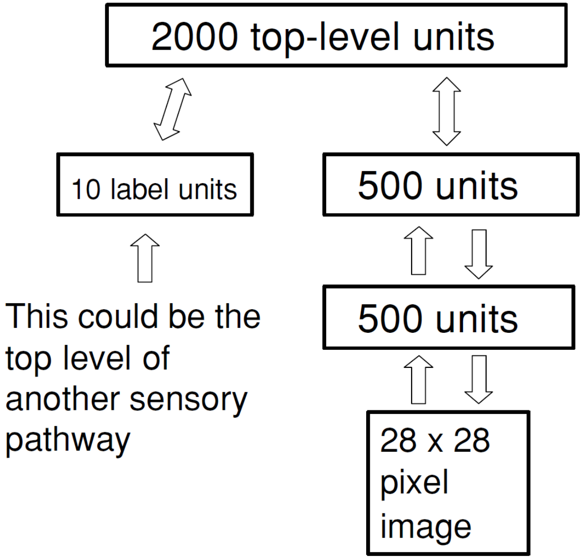
Hinton's 2006…

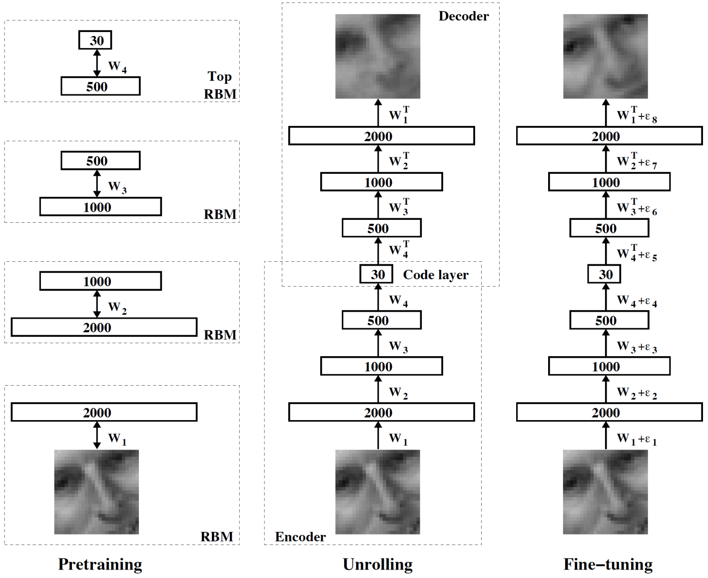
Basic papers on deep learning

Hinton, G. E., Osindero, S. and Teh, Y. (2006)
A fast learning algorithm for deep belief nets.
Neural Computation, **18**, pp 1527-1554. [pdf]
Movies of the neural network generating and recognizing digits

Hinton, G. E. and Salakhutdinov, R. R. (2006)
Reducing the dimensionality of data with neural networks.
Science, Vol. 313. no. 5786, pp. 504 - 507, 28 July 2006.
[full paper] [supporting online material (pdf)] [Matlab code]

DBNs





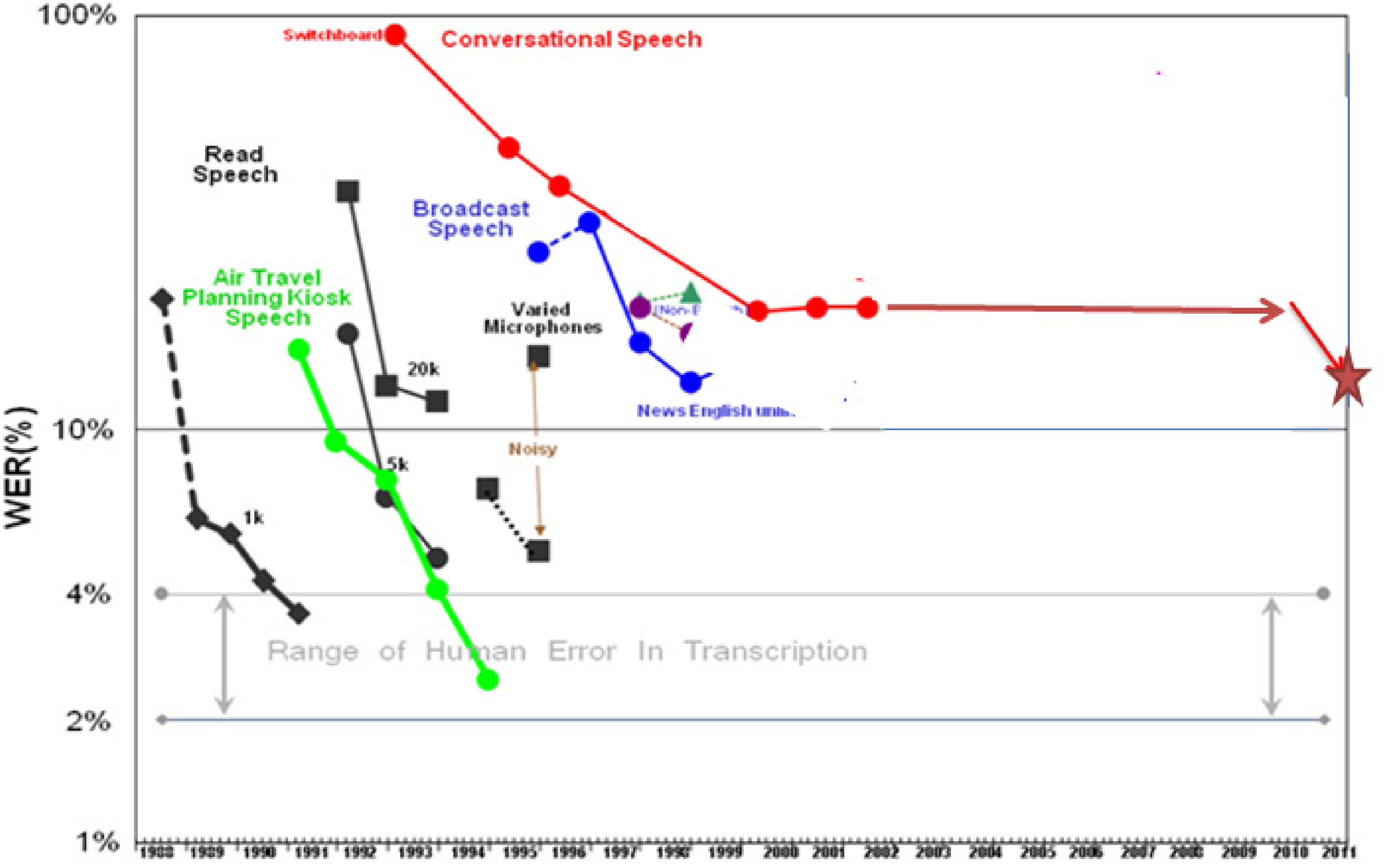
- A DBN is composed of a stack of Restricted Boltzmann machines (RBMs).
- A core component of the DBN is a greedy, layer-by-layer learning algorithm which optimizes DBN weights **at time complexity linear to the size and depth of the networks.**

After 2007…

- 2010年，美国国防部DARPA计划首次资助深度学习项目，参与方包括：斯坦福大学、纽约大学和NEC美国研究院.
- 2012年，Andrew Ng与Jeff Dean搭建Google Brain项目，用包含16000个CPU核的并行计算平台训练超过10亿个神经元的深度神经网络，在语音识别和图像识别等领域取得突破性进展；
- 2013年，Hinton创立的DNN Research公司被Google收购；
- 2013年，Yann LeCun加盟Facebook的人工智能实验室；
- 2014年，Andrew Ng加盟百度深度学习研究院；
- 2015年，Bengio创立深度学习技术孵化器Element AI;

After 2007…

- 2007年，微软开始与Hinton合作，将DNN用于语音识别；
- 2011年，微软研究院和谷歌的语音识别研究人员先后采用DNN技术降低语音识别错误率20%-30%，是该领域10年来最大突破；
- 2015 年，IBM Watson 的英语会话语音识别系统在Switchboard 数据库中取得了 8% 的词错率（WER）；2016年 5 月，词错率降低至6.9；
- 2016 年 9 月微软新系统在Switchboard 语音识别任务上，取得了6.3% 的词错率；
- 2015 年 12 月百度的Deep Speech 2 发布时，首席科学家吴恩达表示其识别的精度已经超过了 Google Speech API、wit.ai、微软的 Bing Speech 和苹果的 Dictation 至少 10 个百分点。到2016年 2 月份时，Deep Speech 2 的短语识别的词错率已经降到了 3.7%

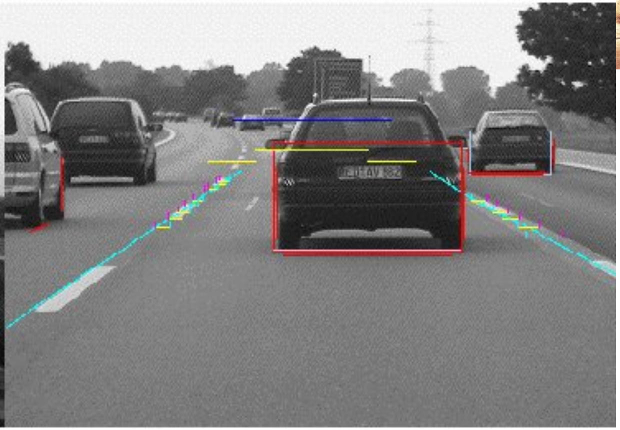

100%
Switchboard
Conversational Speech
Read
Speech
Broadcast
Speech
Air Travel
Planning Kiosk
Speech
Varied
Microphones
20k
5k
1k
Noisy
News English
10%
WER(%)
4%
2%
1%
Range of Human Error In Transcription

1989 2011

After 2007…

- 2009年，Google开始与Hinton开展合作研究；
- 图像分类：
  - 2012年的ImageNet大规模图像识别竞赛（ILSVRC2012）中以超过第二名10个百分点的成绩(83.6%的Top5精度)碾压第二名（74.2%，使用传统的计算机视觉方法）
  - 2013年ImageNet 大规模图像识别竞赛冠军的88.8%
  - 2014年VGG的92.7%和同年的GoogLeNet的93.3%
  - 2015年，在1000类的图像识别中，微软提出的ResNet以96.43%的Top5正确率，超过人类的平均水平；人类的正确率也只有94.9%.
- 物体识别
  - 在PASCAL VOC数据集上的检测平均精度（mAP）：R-CNN（53.3%），Fast RCNN（68.4%），Faster R-CNN的（75.9%），Faster RCNN+Resnet-101，精度可以达83.8%。
  - 检测速度：RCNN，处理一张图片2秒多；Faster RCNN达198毫秒/张，YOLO达155帧/秒（精度52.7%），SSD达23帧/秒（精度75.1%）

Computer Vision



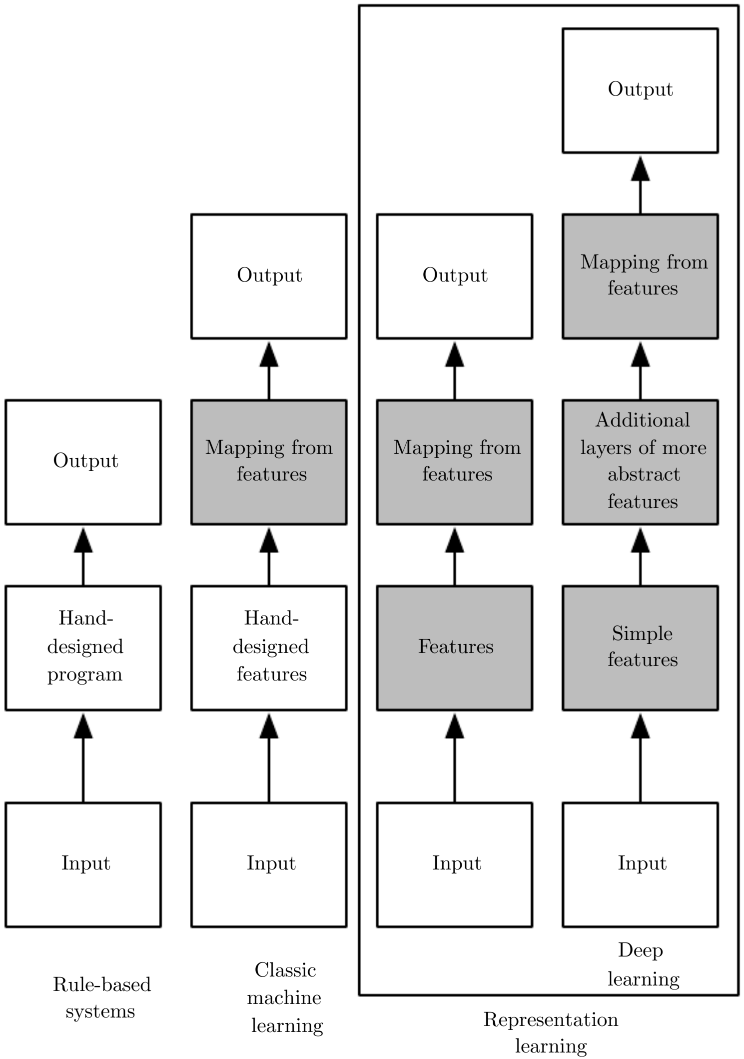
After 2007…

- 2013年,神经机器翻译（Neural Machine Translation, NMT）采用神经网络直接实现两个句子之间的自动翻译，完全没有规则方法和 SMT 方法的从小片段组装成大片段翻译的过程。
- 2014年，Cho 和 Sutskever 提出了 Encoder-Decoder 架构的神经网络机器翻译系统。
- 2015年，Yoshua Bengio 团队进一步加入了Attention 的概念。Bengio 团队的这个工作也奠定了后续很多NMT商业系统的基础.
- 2016年9月，Google发布神经机器翻译系统（GNMT），实现了机器翻译领域的重大突破。
- 2016年11月，Google发论文宣布了其在多语言机器翻译上实现了zero-shot翻译。

After 2007…

- 2013年，百度成立“深度学习研究所”（IDL），将深度学习应用于语音识别和图像识别、检索，以及广告CTR预估（Click-Through-Rate Prediction，pCTR），其中图片检索达到了国际领先水平。2014年，Andrew Ng加盟。
- 2012年，华为在香港成立“诺亚方舟实验室”从事自然语言处理、数据挖掘与机器学习、媒体社交、人机交互等方面的研究。2014年
- 2013年，腾讯着手建立深度学习平台Mariana。Mariana面向语音识别、图像识别、广告推荐等众多应用领域，提供默认算法的并行实现。
- 2015年，阿里发布DTPAI人工智能平台，称其中包含深度学习的开放模块；

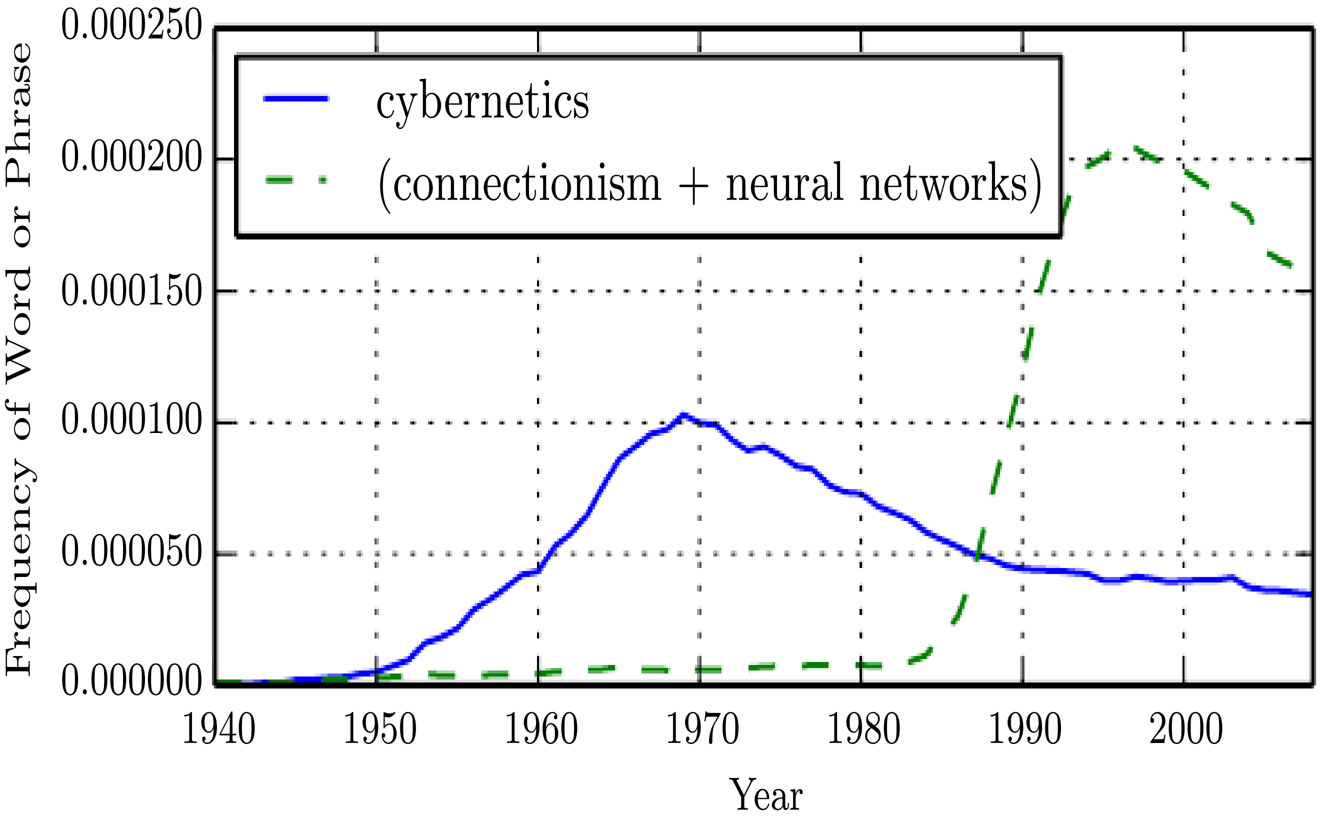
Key Trends about Deep Learning



- Flowcharts showing how the different parts of an AI system relate to each other within different AI disciplines. Shaded boxes indicate components that are able to learn from data.

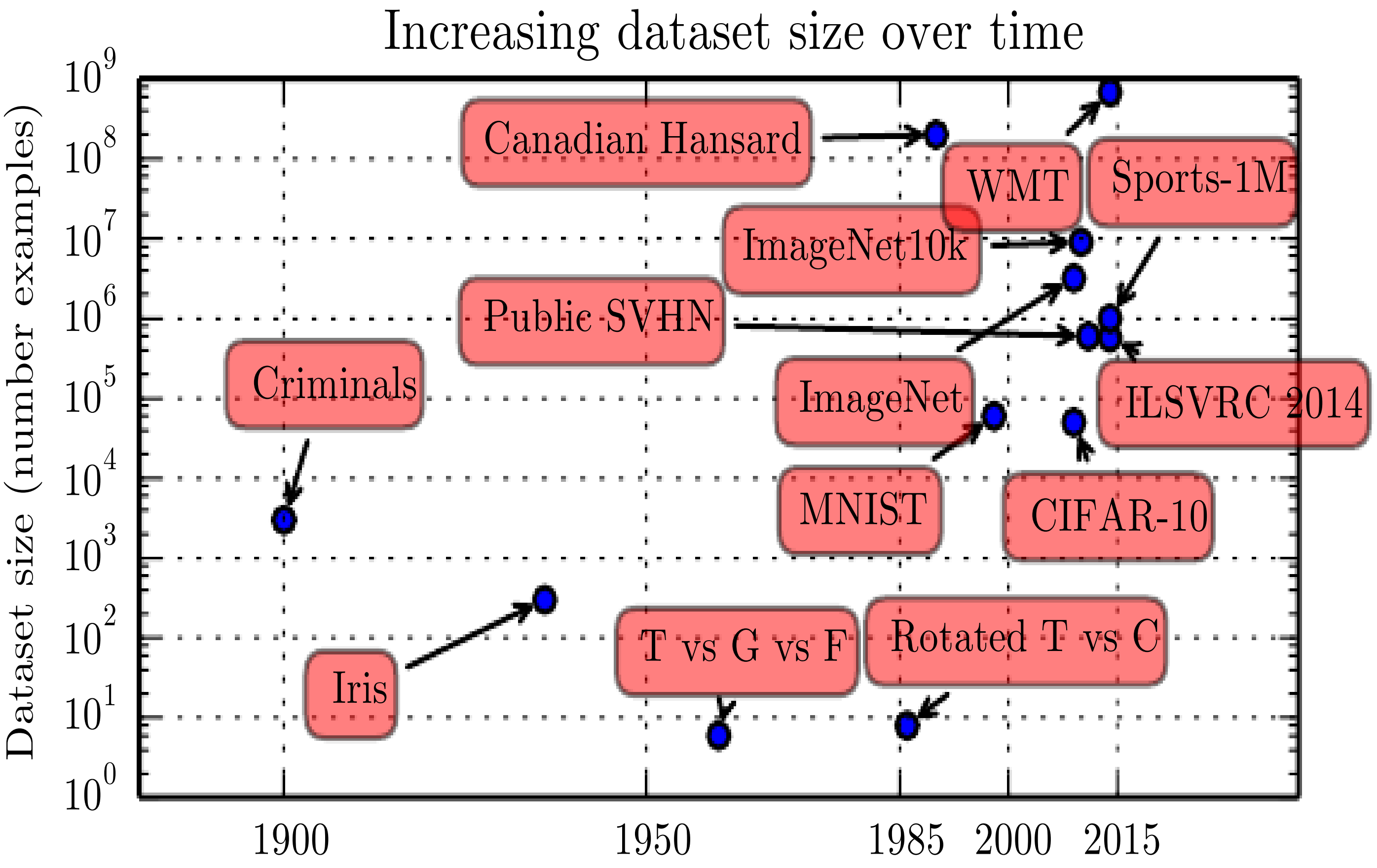
Key Trends about Deep Learning

- The figure shows two of the three historical waves of artificial neural nets research, as measured by the frequency of the phrases “cybernetics” and “connectionism” or “neural networks” according to Google Books

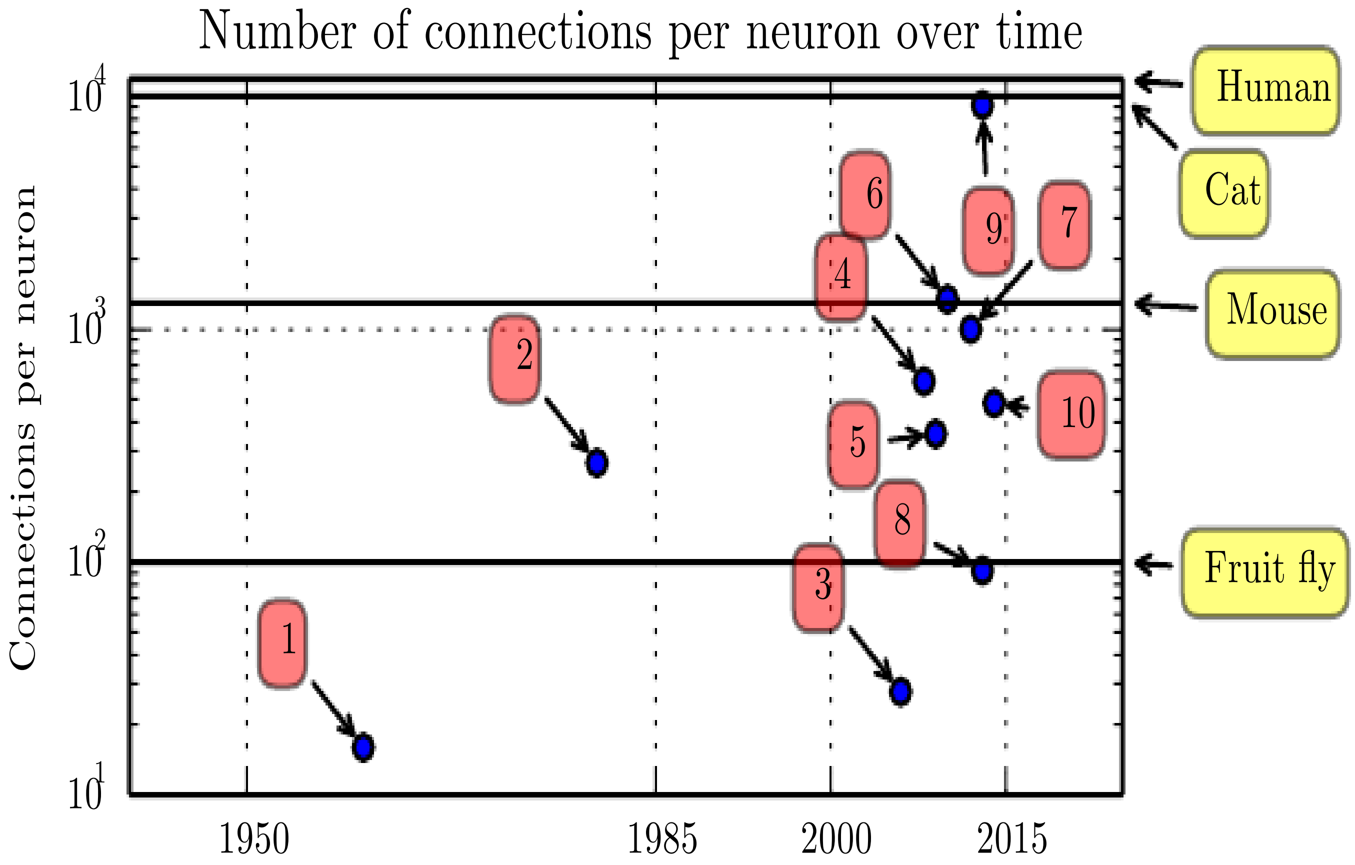


Key Trends about Deep Learning

1. Deep learning has become more useful as the amount of available training data has increased.



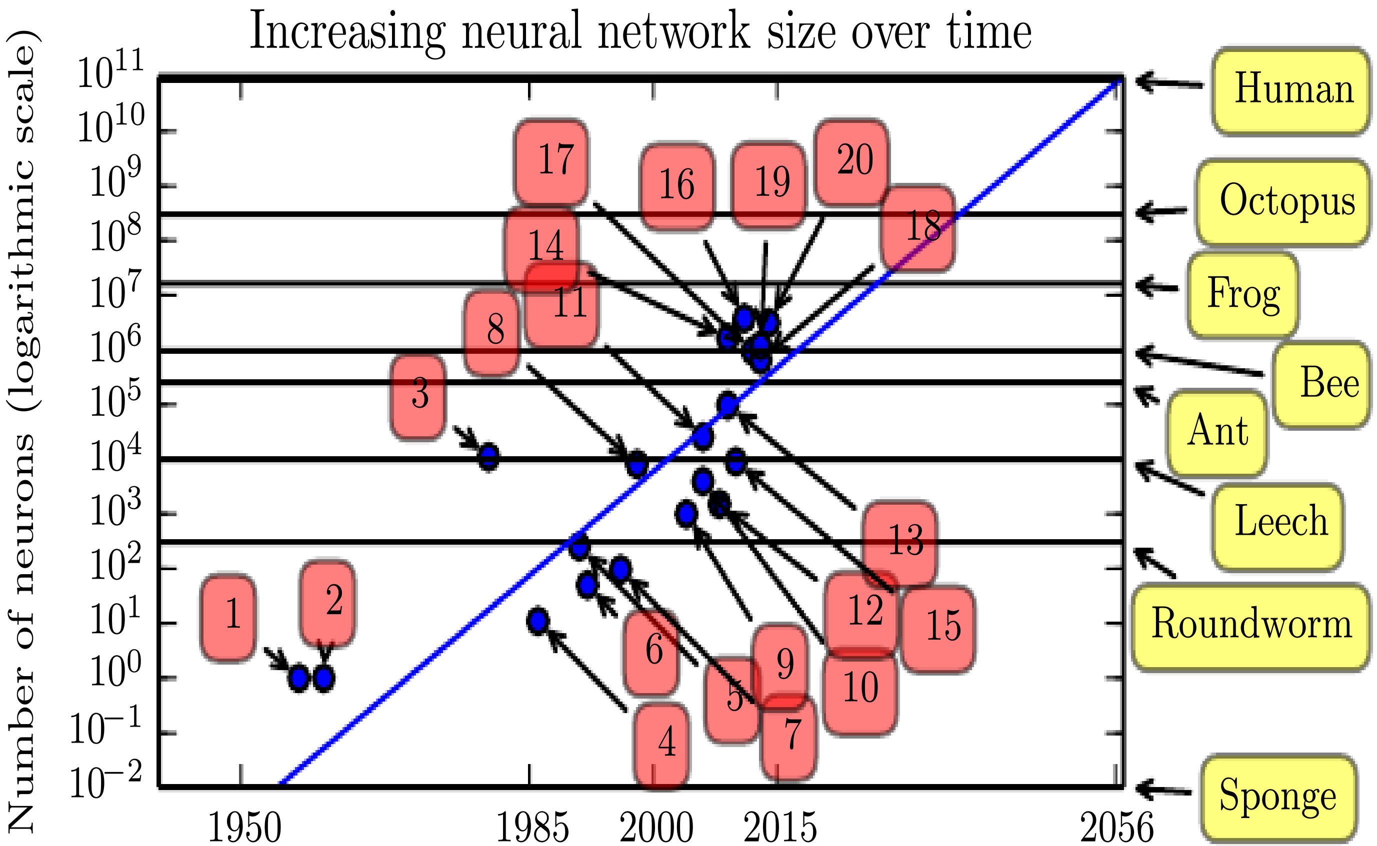
Key Trends about Deep Learning



2. Deep learning models have grown in size over time as computer hardware and software infrastructure for deep learning has improved.

Key Trends about Deep Learning

2. Deep learning models have grown in size over time as computer hardware and software infrastructure for deep learning has improved.



Challenge to AI

- The true challenge to artificial intelligence
  - proved to be solving the tasks that are **easy for people to perform but hard for people to describe formally**;
  - solving the problems that we solve intuitively, that feel automatic;
  - like recognizing spoken words or faces in images;

当前AI面临的技术问题

低效

- **像工业革命前的“蒸汽机原型”一样，低效而昂贵；**
- 依赖大量“数据矿石”作为燃料，抵消机器本身的缺陷;
- 谷歌、Facebook、微软用海量数据应对低效的机器学习系统，小企业没有数据，难以实现；

不通用

- 通常需要“针对特定任务建立特定的智能系统”**难以建立具有“通用智能”的系统；**
- Google Deep Mind：“还没有人工智能可以同时学会多款不同游戏的玩法” “甚至无法让机器学会不同游戏”；

不透明

- 很多AI技术对观察者来说是“黑盒”，尽管可以看到输入与输出，但其内部决策机制通常无法解释；
- **“知其然，不知其所以然”**导致很多AI系统的训练与调整变得昂贵、低效、缺乏科学方法指导；

Thanks.