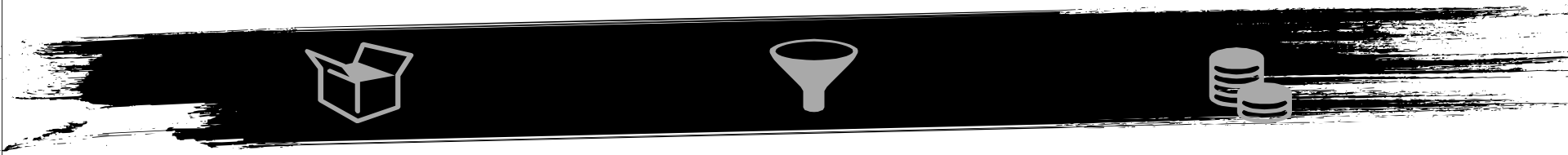
Shaping up Internet Search with Deep Learning

Kai Yu

Baidu IDL

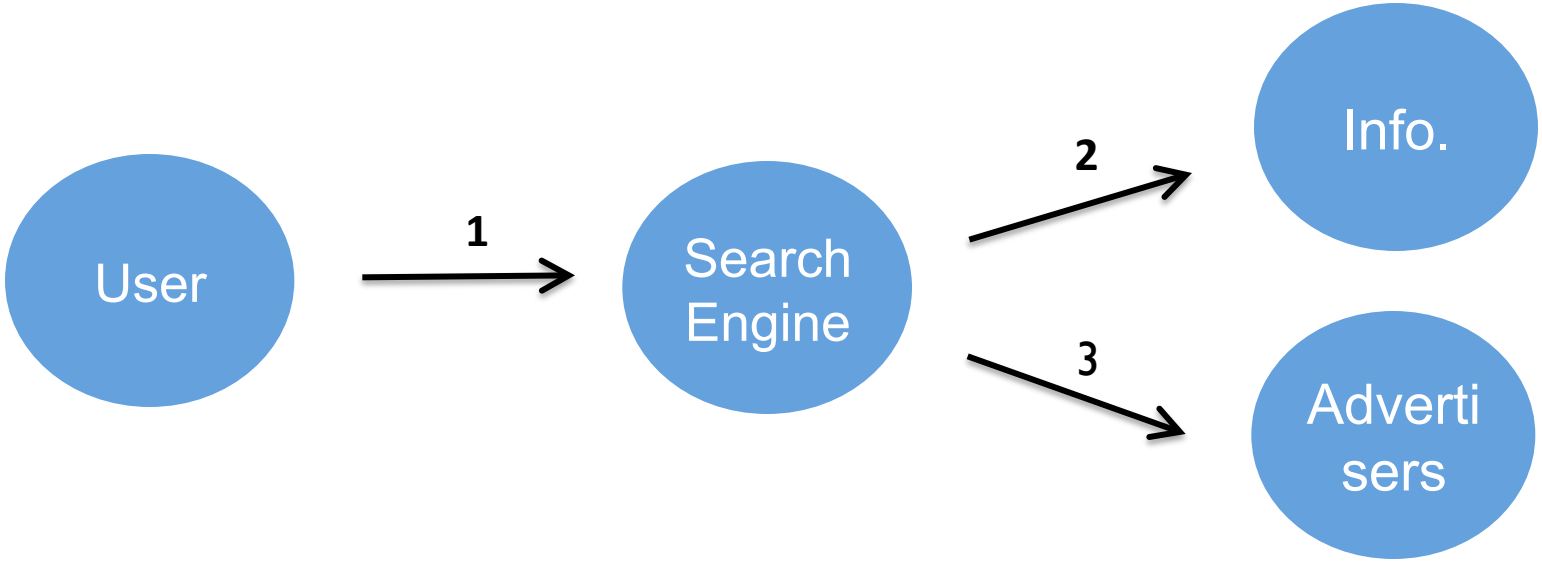
Our business model in a nutshell



Mega Data | Data Mining | Monetization

Turn data into value via technology

Search engine and machine learning

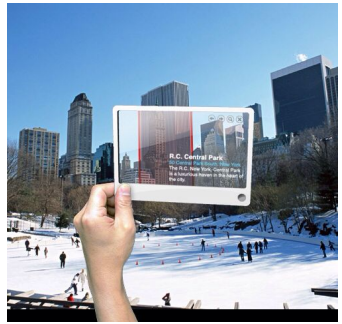


1. Query understanding
2. Search ranking
3. CTR estimation

Search is evolving

PC Search

Mobile Search



Future Search

- Natural human-computer interface
- Semantic understanding of contents

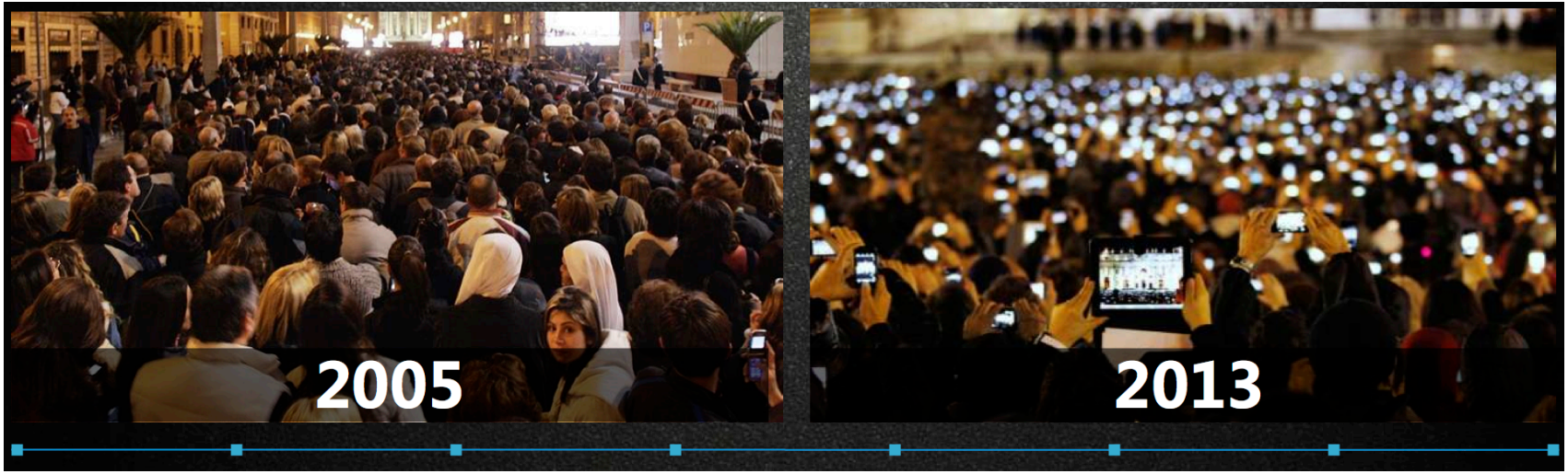
9 technology challenges from Baidu

On Aug 13, 2012, CEO Robin Li gave a keynote speech at ACM KDD, and proposed 9 major technology challenges to the academic research community. The first 3 are:

1. OCR in natural images
2. Speech recognition and understanding
3. Content-based image retrieval (visual search)

Visual sensing by mobile phones

Visual sensing by mobile phones



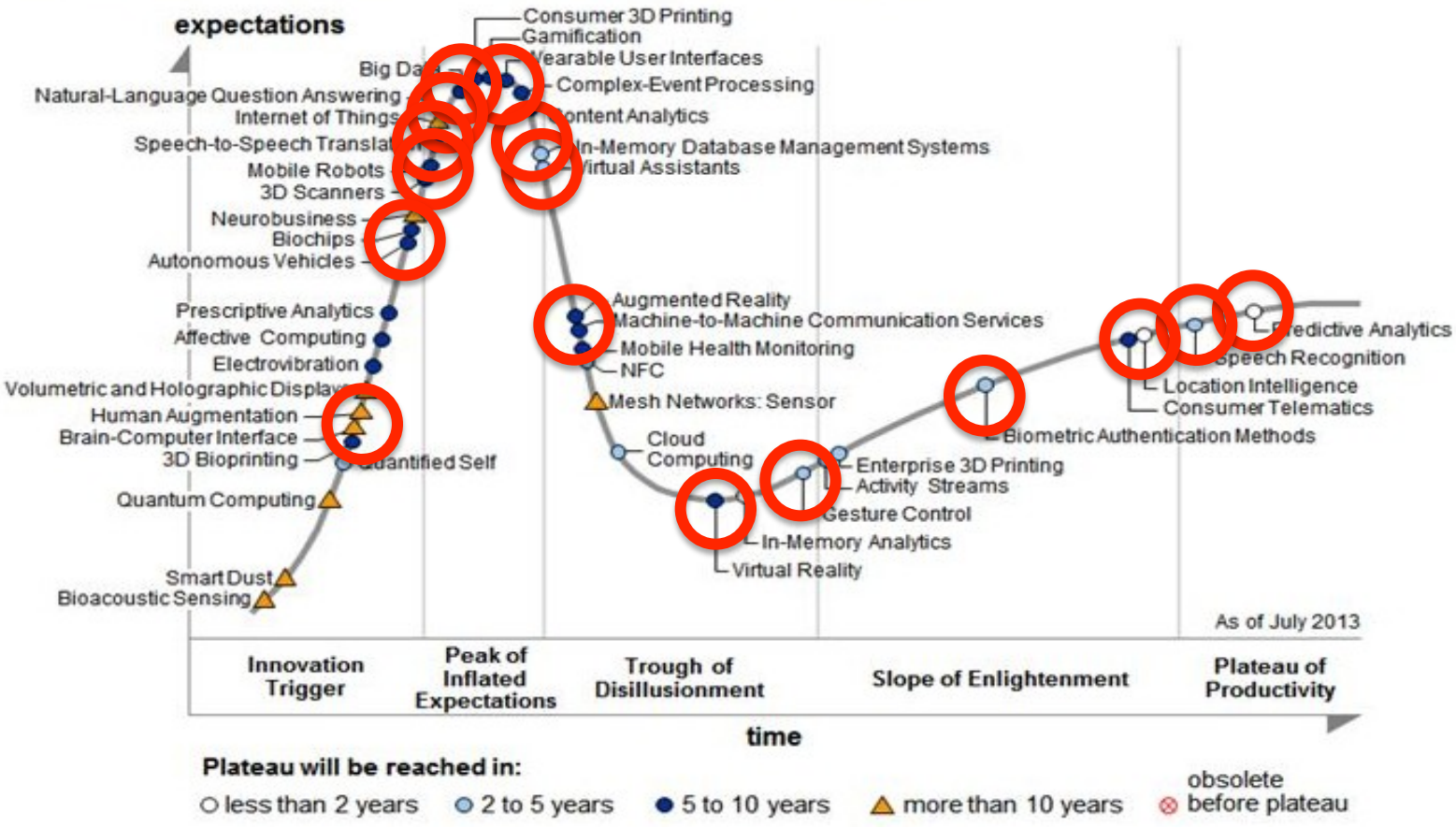
A global race on speech recognition

NUANCE®

iFLY

Baidu 百度

Gartner Emerging Tech Hype Cycle 2013



Machine Learning

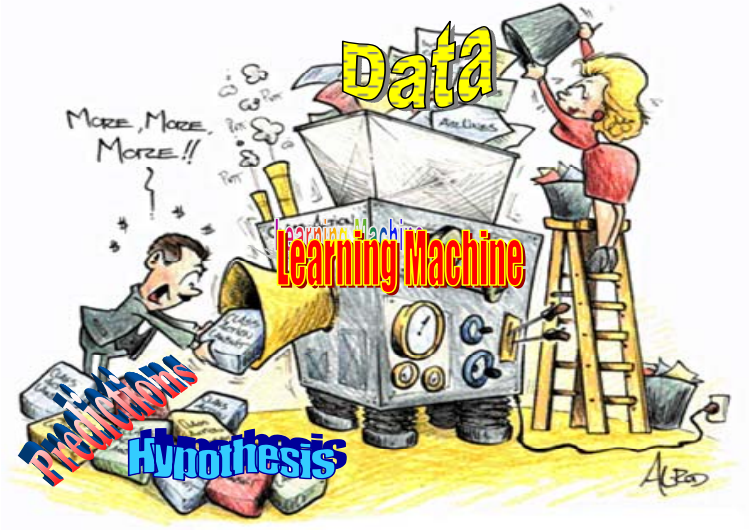
Machine Learning

Given observations (X_i, Y_i), $i = 1, \ldots, n$

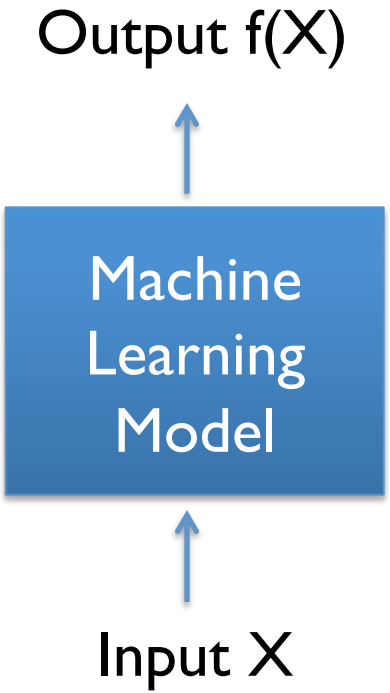
Learn a predictive function $f(X)$

Generalization error $E[L(f(X), Y)]$

Empirical loss $\Sigma L(f(X_i), Y_i) / n$



What can machine learning be used for



- 输入语音，输出文字
- 输入一物体的图像，输出该物体名称
- 输入用户购物历史，输出其潜在需求
- 输入过去股票价格，输出明天的价格

Generalization error decomposition

$$E[\, L(\, f(X), Y\,)\,] = A + E$$

- Approximation error – model class
- Estimation error – data size

Generalization error decomposition

$$E[\, L(\, f(X), Y\,)\,] = A + E + O$$

- Approximation error – model class
- Estimation error – data size
- Optimization error – algorithm

Generalization error decomposition

$$E[\, L(\, f(X), Y\,)\,] = A + E + O$$

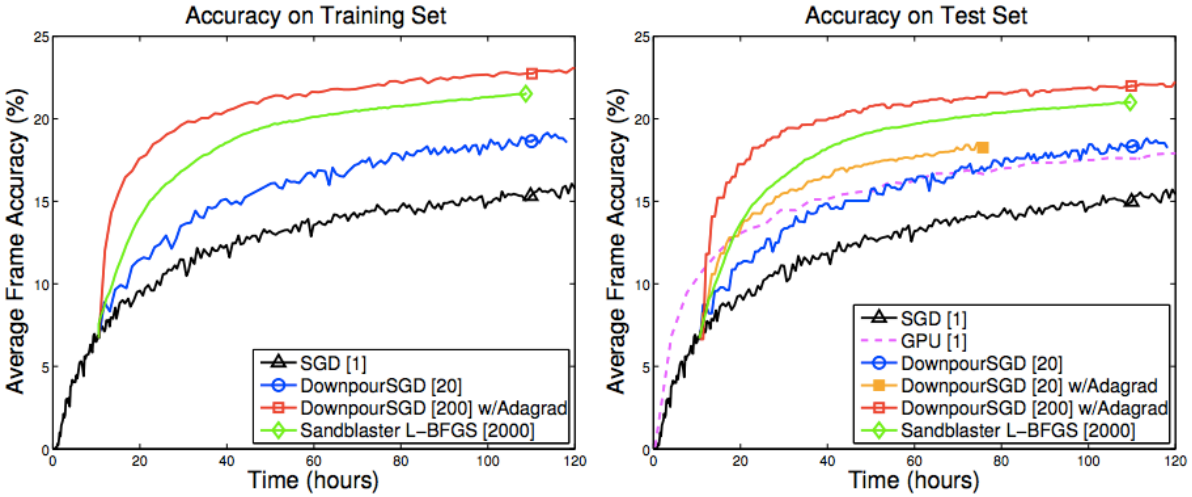
- Approximation error – model class : use complex model
- Estimation error – data size : collect big data
- Optimization error – algorithm : design optimization algorithm

Generalization error decomposition

$$E[\, L(\, f(X), Y\,)\,] = A + E + O$$

- Approximation error – model class : use complex model
- Estimation error – data size : collect big data
- Optimization error – algorithm : design an OK algorithm

Why do we need complex models



Observed underfitting on speech data

Large-scale distributed deep networks, Jeff Dean et al, NIPS 12

Deep Learning

Deep Learning Since 2006

materials are identical for all configurations. The blue bars in Fig. 1 summarize the measured SHG signals. For excitation of the *LC* resonance in Fig. 1A (horizontal incident polarization), we find an SHG signal that is 500 times above the noise level. As expected for SHG, this signal closely scales with the square of the incident power (Fig. 2A). The polarization of the SHG emission is nearly vertical (Fig. 2B). The small angle with respect to the vertical is due to deviations from perfect mirror symmetry of the SRRs (see electron micrographs in Fig. 1). Small detuning of the *LC* resonance toward smaller wavelength (i.e., to 1.3-μm wavelength) reduces the SHG signal strength from 100% to 20%. For excitation of the Mie resonance with vertical incident polarization in Fig. 1D, we find a small signal just above the noise level. For excitation of the Mie resonance with horizontal incident polarization in Fig. 1C, a small but significant SHG emission is found, which is again po-

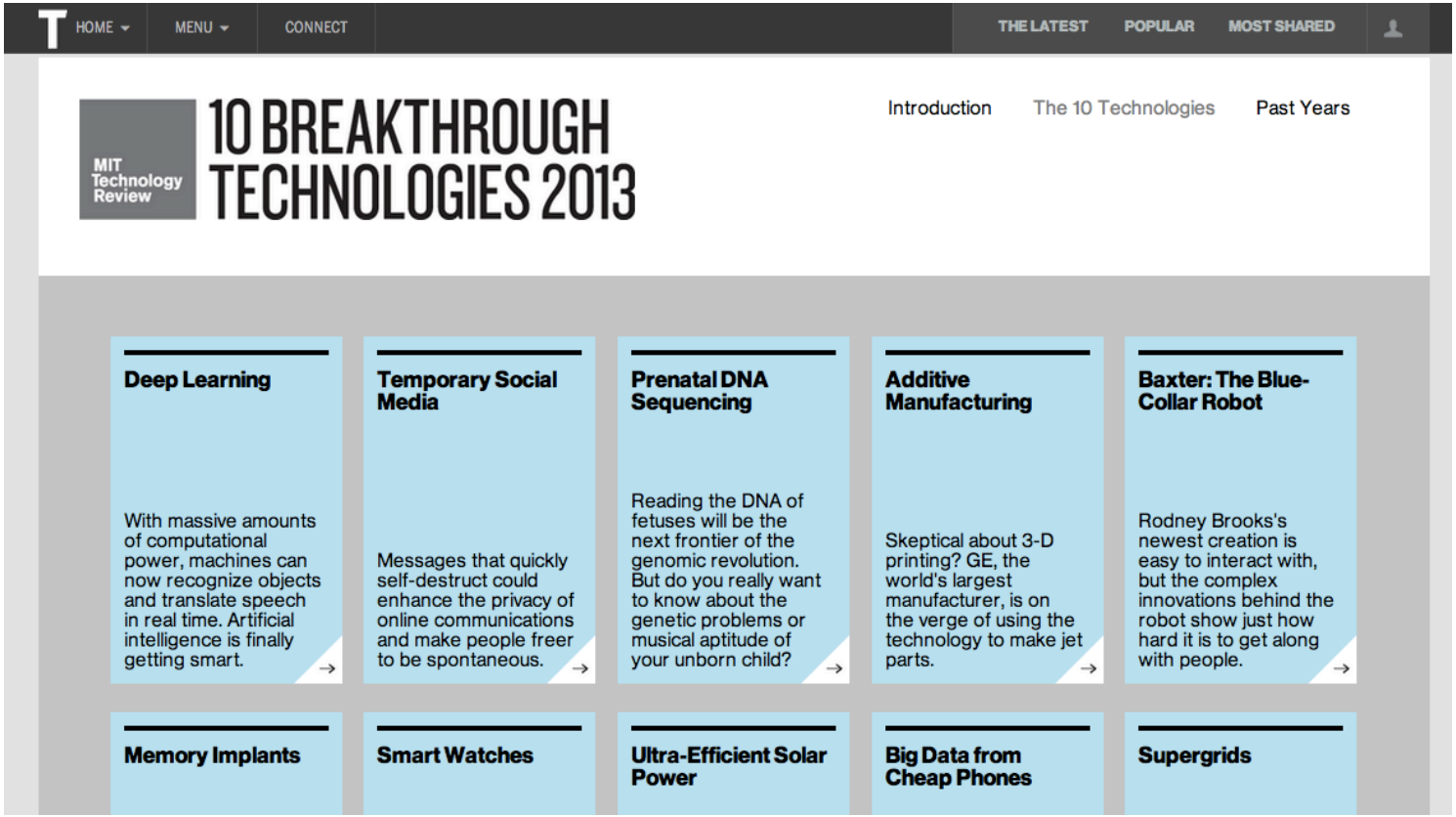
Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton* and R. R. Salakhutdinov

High-dimensional data can be converted to low-dimensional codes by training a multilayer neural network with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" networks, but this works well only if the initial weights are close to a good solution. We describe an effective way of initializing the weights that allows deep autoencoder networks to learn low-dimensional codes that work much better than principal components analysis as a tool to reduce the dimensionality of data.

Dimensionality reduction facilitates the classification, visualization, communication, and storage of high-dimensional data. A simple and widely used method is principal components analysis (PCA), which finds the directions of greatest variance in the data set and represents each data point by its coordinates along each of these directions. We describe a nonlinear generalization of PCA that uses an adaptive, multilayer "encoder" network

Top breakthrough technology 2013



MIT Technology Review, April 23rd, 2013

Revolution on Speech Recognition

task	hours of training data	DNN-HMM	GMM-HMM with same data
Switchboard (test set 1)	309	18.5	27.4
Switchboard (test set 2)	309	16.1	23.6
English Broadcast News	50	17.5	18.8
Bing Voice Search (Sentence error rates)	24	30.4	36.2
Google Voice Input	5,870	12.3	
Youtube	1,400	47.6	52.3

Slide Courtesy: Geoff Hinton

图像识别领域的突破

ImageNet Challenge

72%, 2010

74%, 2011

85%, 2012

Deep Learning in Industry

Scientists See Promise in Deep-Learning Programs

A voice recognition program translated a speec
Chinese.

By JOHN MARKOFF
Published: November 23, 2012

The New York Times

Facebook announced its AI Lab in 2013

COMPUTING NEWS

23 COMMENTS

Facebook Launches Advanced AI Effort to Find Meaning in Your Posts

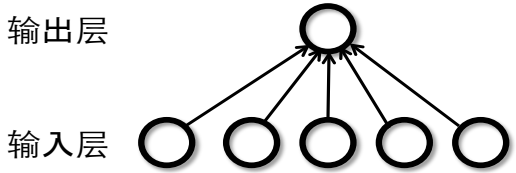
A technique called deep learning could help Facebook understand their data better.

By Tom Simonite on September 20, 2013

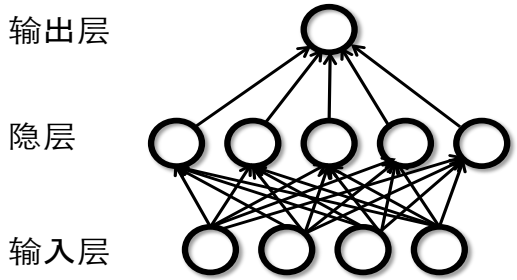
Baidu's commitment to research

- Jan. 2013, announced to build its research lab
- Institute of Deep Learning (IDL)
- The focus is Artificial Intelligence

All Machine Learning Models in One Page

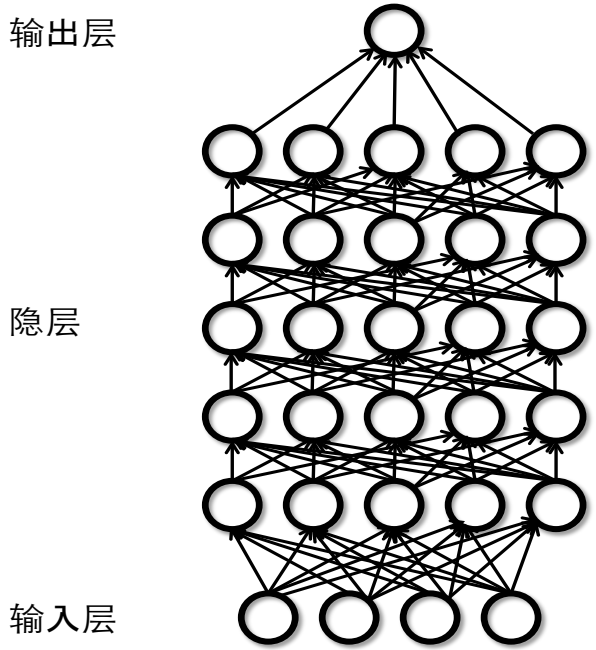


不含隐层的浅层学习模型



含有单个隐层的浅层学习模型

Shallow Models



含多个隐层的深度学习模型

Deep Models

Shallow Models Since Late 80's

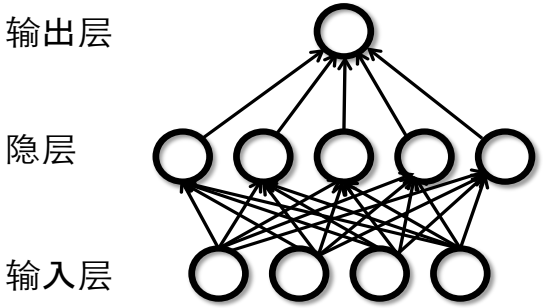
- Neural Networks
- Boosting
- Support Vector Machines
- Maximum Entropy
- …

?

Given good features, how to do classification?

Since 2000 – Learning Hidden Structures

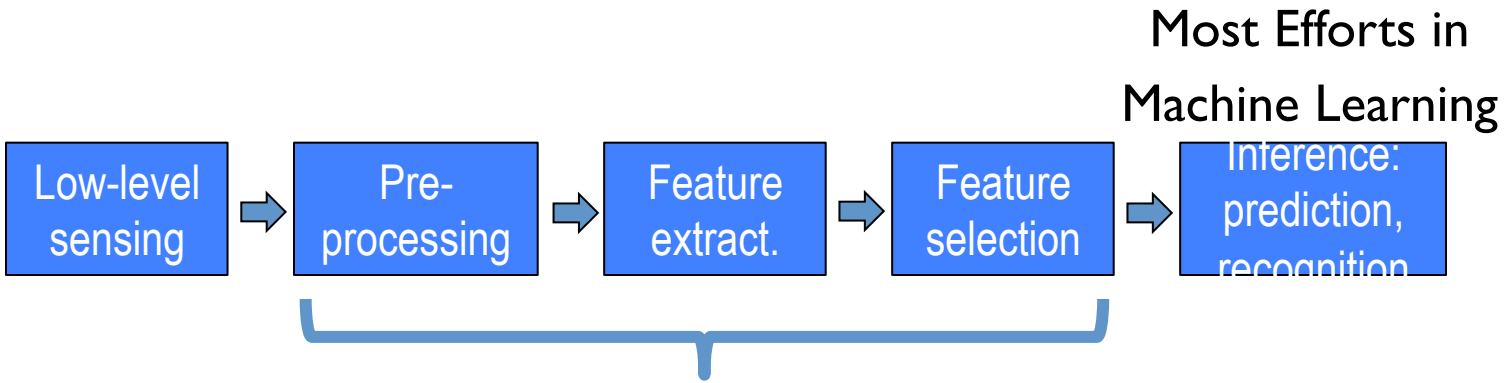
- Kernel Learning
- Transfer Learning
- Semi-supervised Learning
- Manifold Learning
- Matrix Factorization
  - PCA, ICA, Topic Model, …
- Sparse Learning
- …



含有单个隐层的浅层学习模型

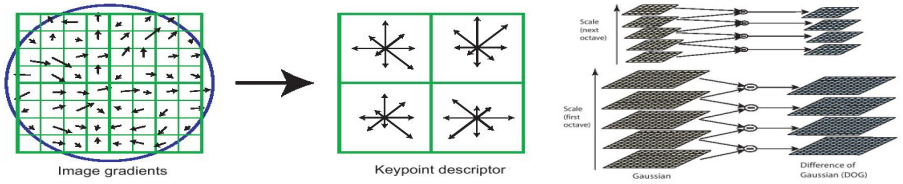
This structure seems to be very universal

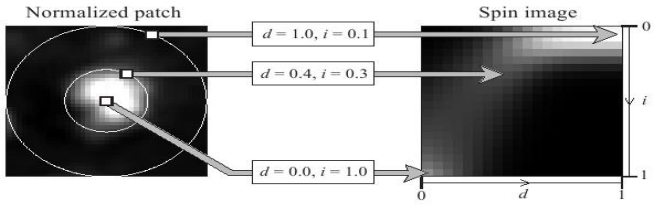
The pipeline of machine visual perception

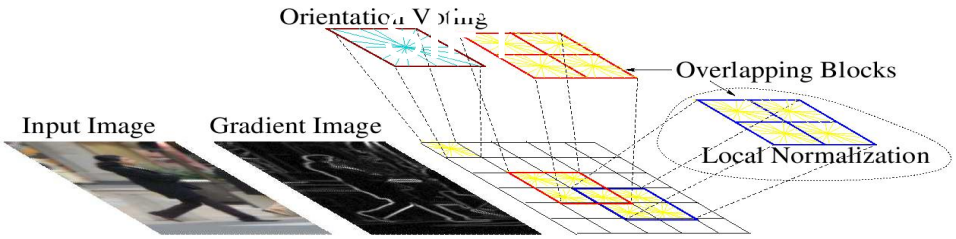


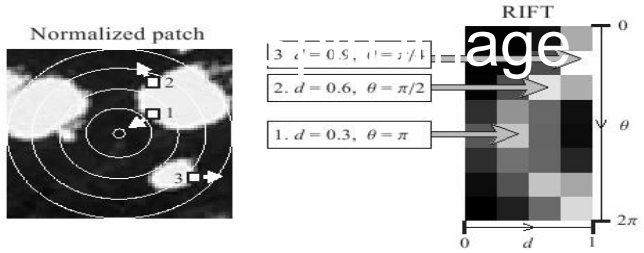
- Most critical for accuracy
- Account for most of the computation for testing
- Most time-consuming in development cycle
- Often hand-craft in practice

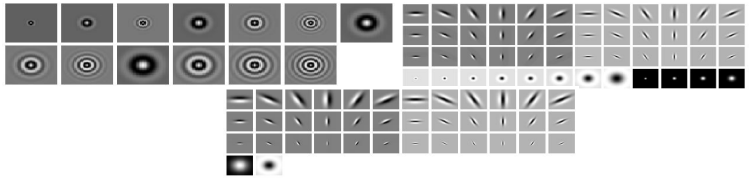
Computer vision features

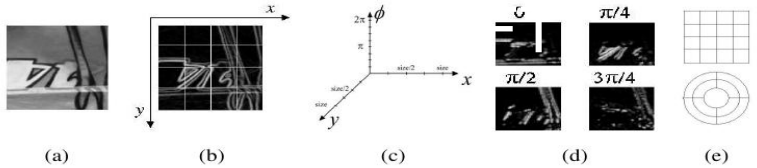






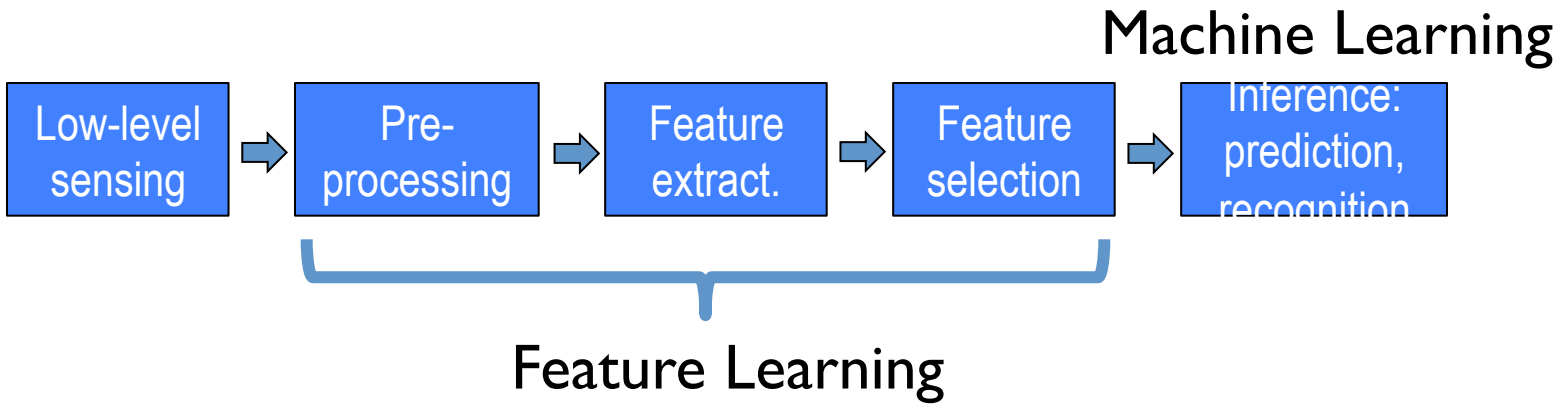




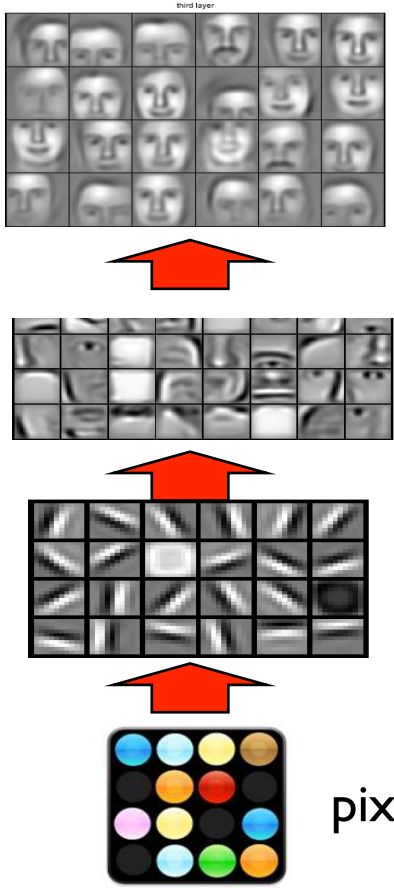


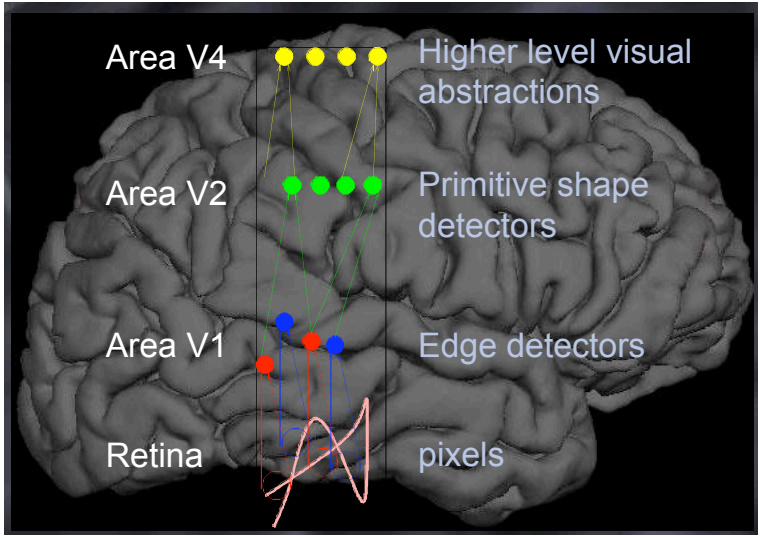
Slide Courtesy: Andrew Ng

Deep Learning: learning features from data



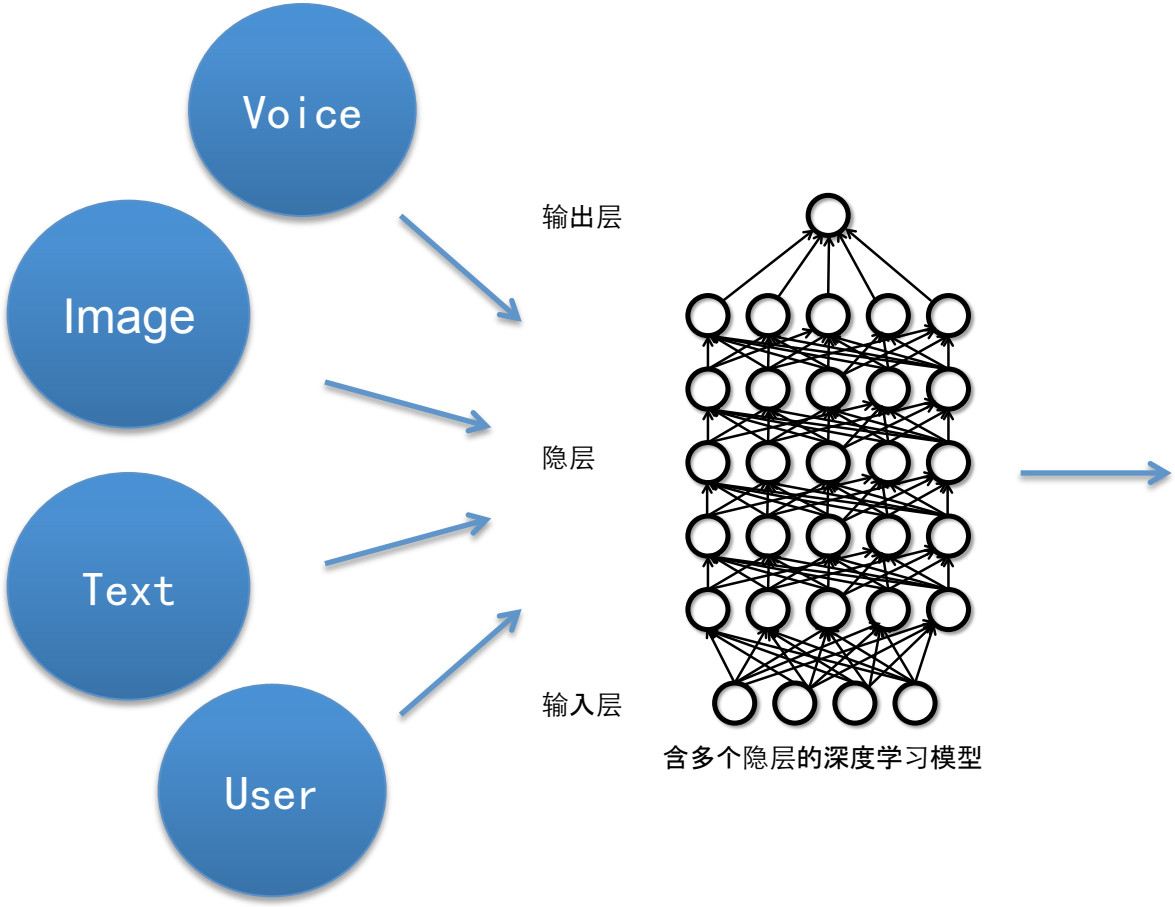
Deep learning vs. the brain





Slide credit: Andrew Ng

Intelligent search powered by DL



Progress of DL at Baidu

- July 2012, get started
- Nov. 2012, big improvements on speech, ocr, face, …
- By 2012 end, 5 DL-based products got online

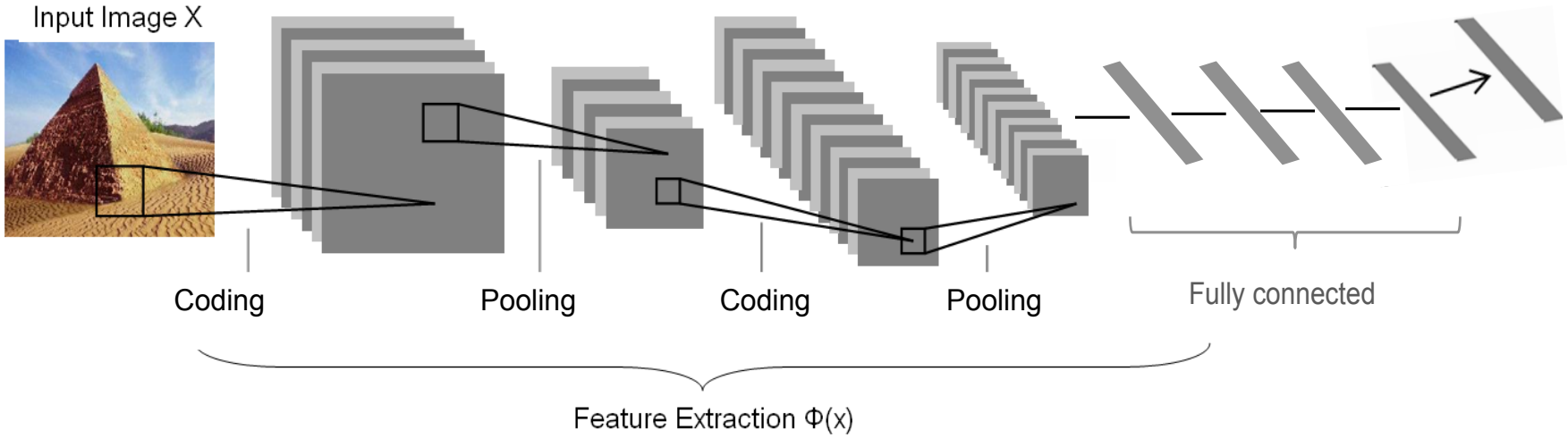
Progress of Deep Learning at Baidu

- Big improvement on speech & image recognition
  - Speech： error rate reduced by 25%
  - OCR： error rate reduced by 30%
  - Image: the best image similarity search system

- Online Ads: DNN CTR for search ads was launched in May 20th 2013, serving billions of search queries everyday – substantial improvement

- Web Search: A DNN semantic model was launched in Dec 2013, which led to the biggest improvement of our search ranking quality.

A deep model for image recognition



A DL model for query-doc relevance

Ranking loss

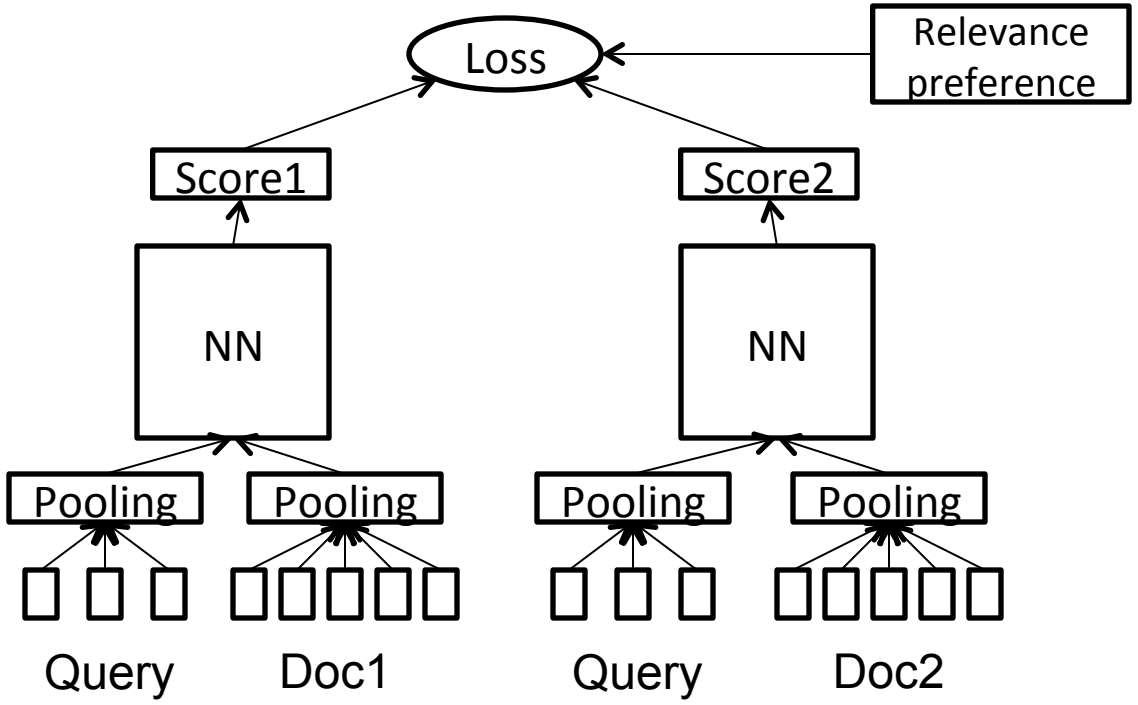
Relevance score

Deep neural network

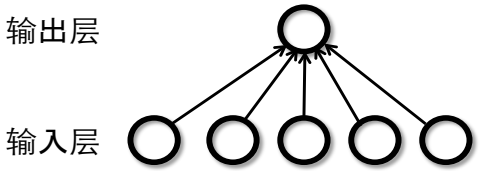
Query and doc representation

Embedding table

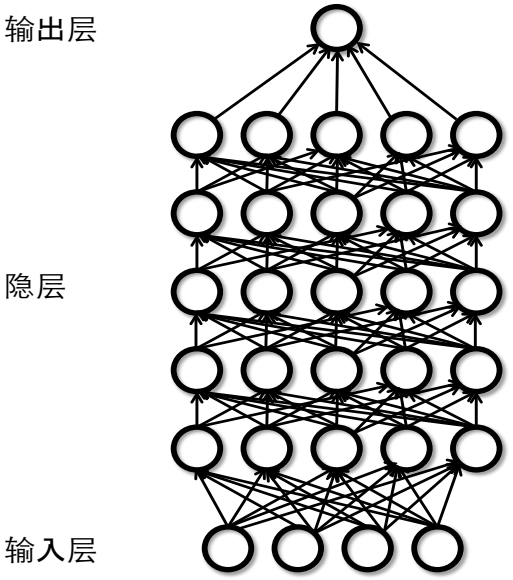
Input



Deep Learning for CTR



不含隐层的浅层学习模型



含多个隐层的深度学习模型

1st generation: shallow models, 100 billion ID features, 100 billion training samples

2nd generation: deep models, features reduced to hundreds dim.,

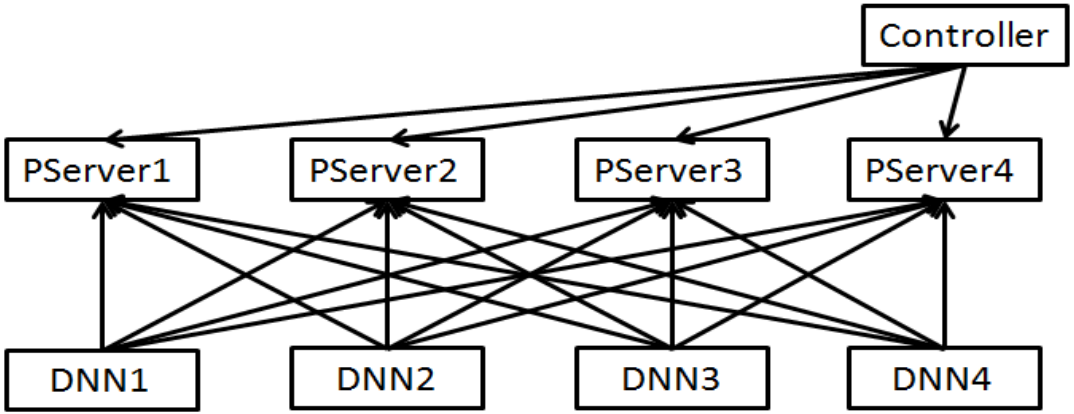
Typical scale of training data at Baidu

- Image recognition: 100 millions
- OCR: 100 millions
- Speech : 10 billions
- CTR: 100 billions
- …

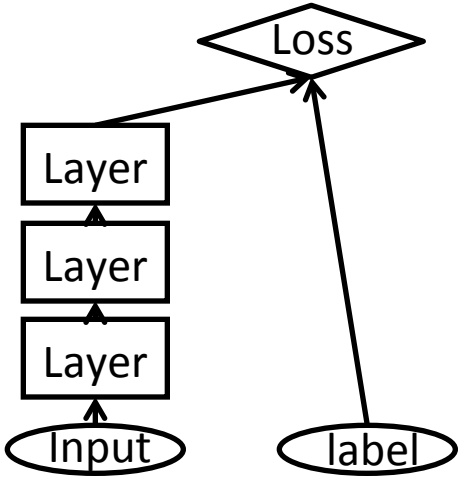
We expect the training data will grow X10 each year

PADDLE Platform

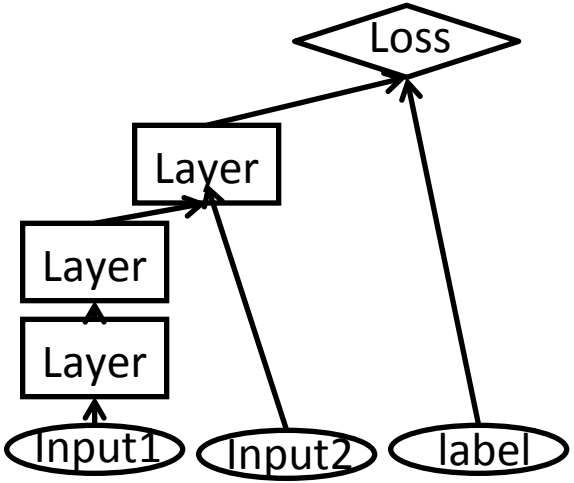
- Use GPUs and CPUs
- Data parallelization.
- Model can be parallelized as well
- Use Parameter Server to coordinate



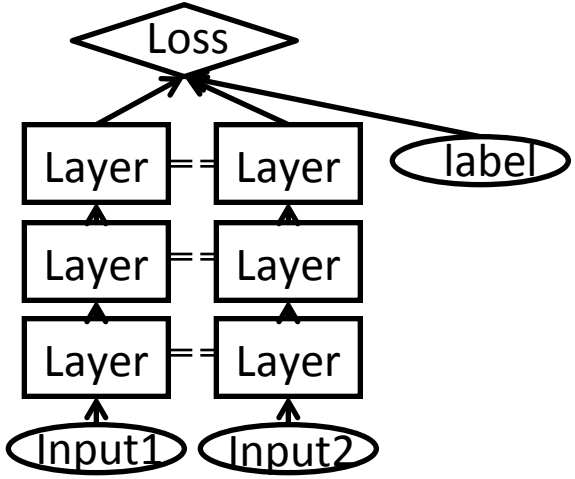
PADDLE: flexible model structures



PADDLE-Speech

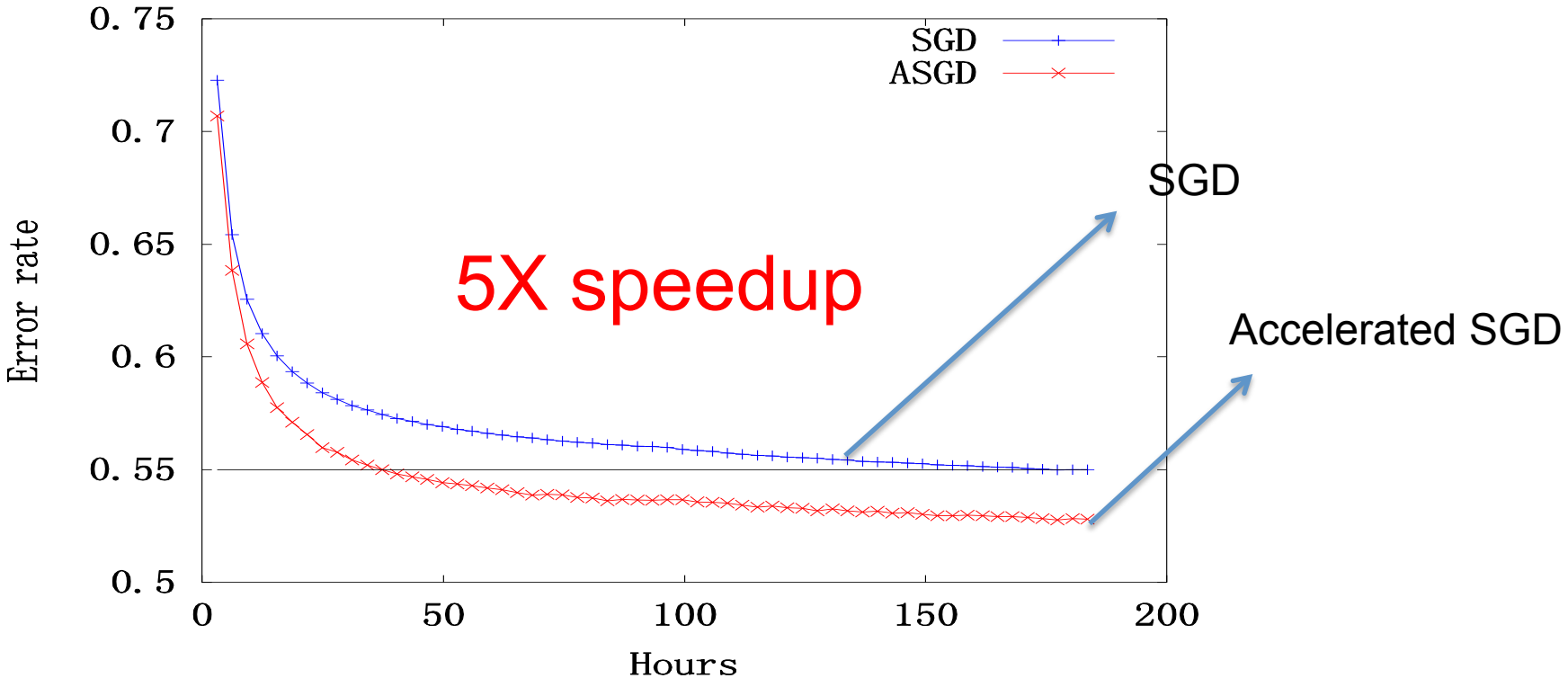


PADDLE-CTR

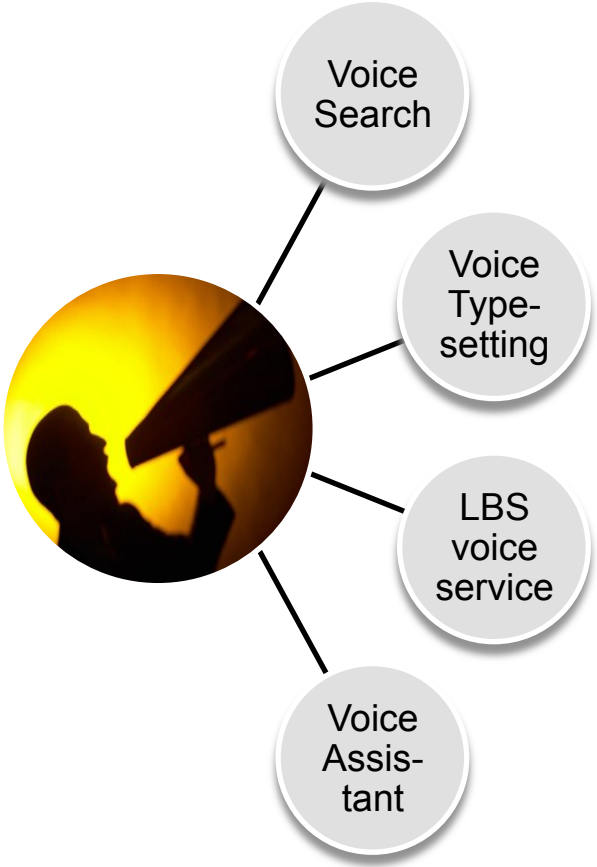


PADDLE-LTR

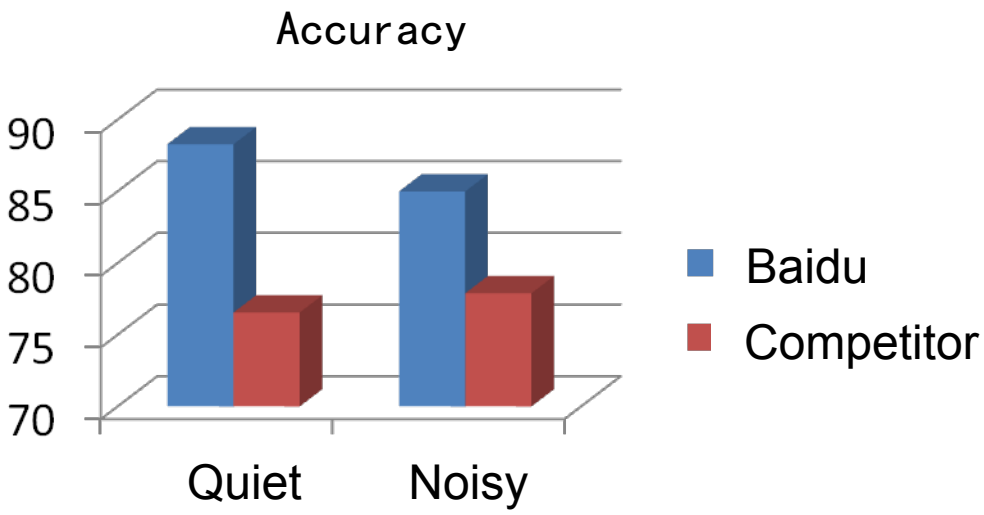
An accelerated SGD algorithm

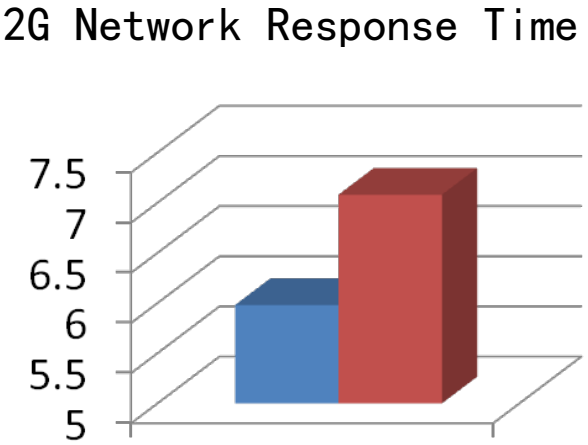


Speech recognition in many products



Recognition accuracy for voice search





For mobile search, the proportion of voice queries has grown to 10 % in the end of 2012.

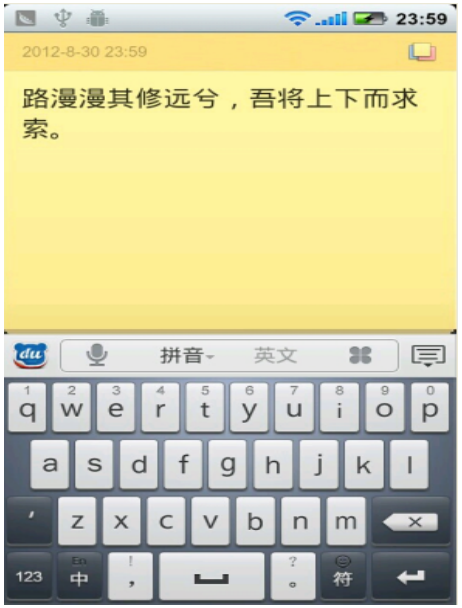
Baidu map voice search

http://shouji.baidu.com/map/

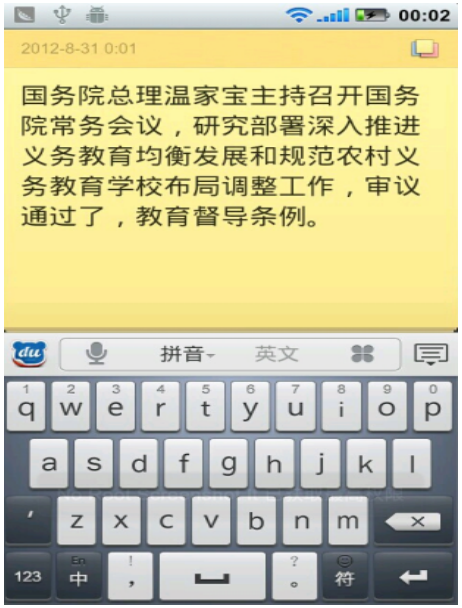
Baidu mobile voice typing

Short message

Weibo



Poem



News reading

http://shouji.baidu.com/input/

Baidu Top Award 2013

For building the best industrial mandarin speech recognition system.

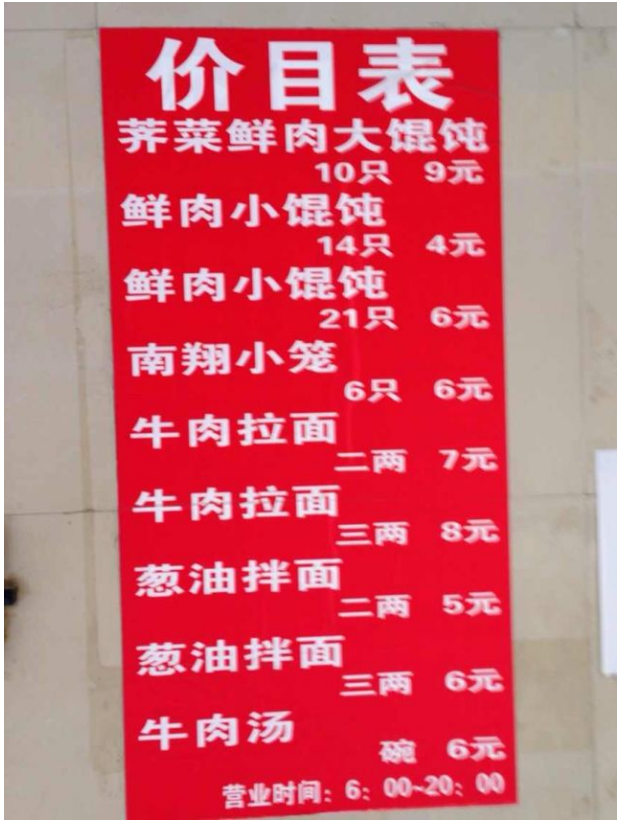
OCR in natural photos

➢OCR search

➢Translation App

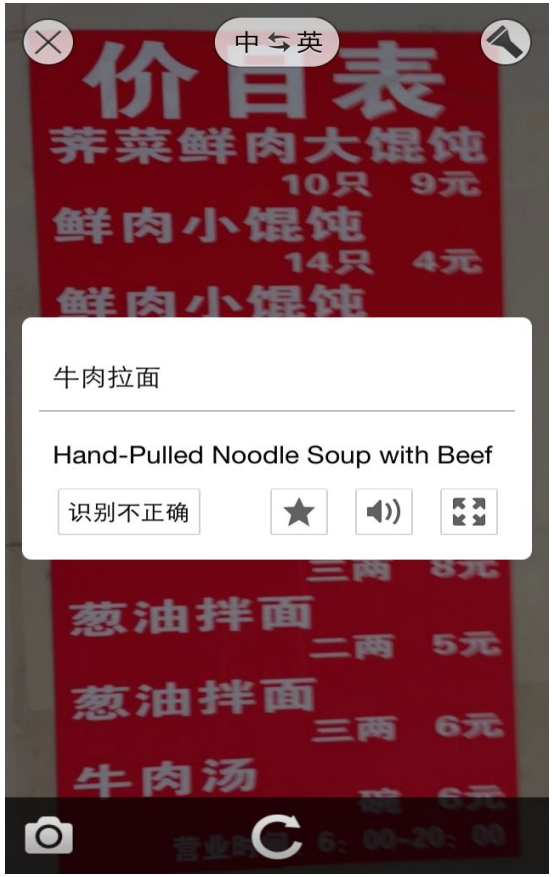
➢Spam detection

OCR in natural photos

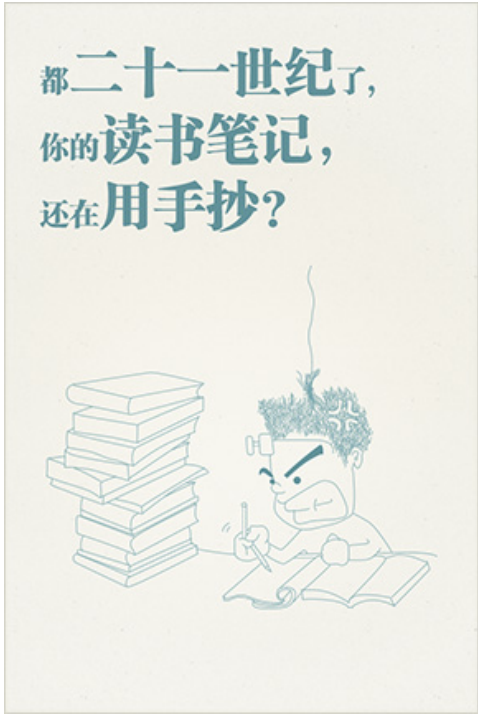


OCR in natural photos

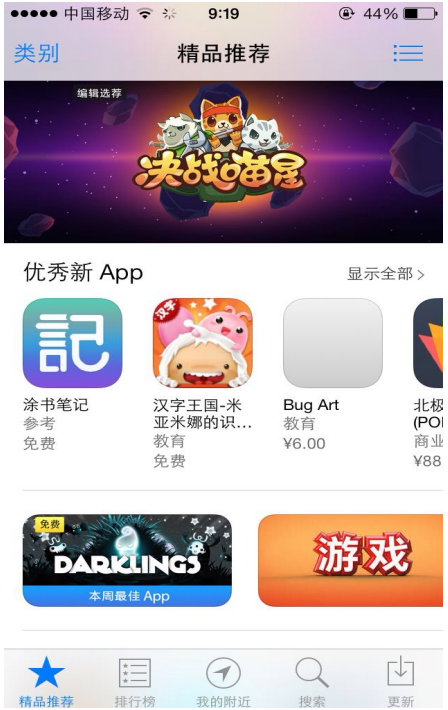
OCR in natural photos



涂书笔记


都二十一世纪了，
你的读书笔记，
还在用手抄？

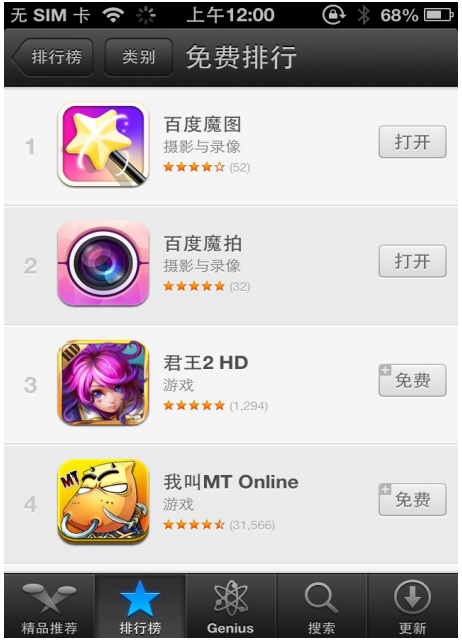

再涂


中国移动
9:19
44%
类别
精品推荐
编辑选荐
决战喵星
优秀新 App
显示全部
涂书笔记
参考
免费
汉字王国-米亚米娜的识…
教育
免费
Bug Art
教育
¥6.00
免费
DARKLINGS
本周最佳 App
游戏
精品推荐
排行榜
我的附近
搜索
更新

Face grouping in online albums

Baidu Photo Wonder (百度魔图)

Baidu Photo Wonder



Users upload 90 million photos one day， ranked TOP on App Store for three weeks

Baidu Photo Wonder

百度魔图

相似度: 74.41%

布死痕象啊~绳命作弄人~(>_<)~

我的照片　　丹尼尔·克雷格

快来下载百度魔图
看看你最像哪位明星吧!

百度魔图

相似度: 85.65%

加把劲儿，马上你也是明星了!

我的照片　　蔡文胜

快来下载百度魔图
看看你最像哪位明星吧!

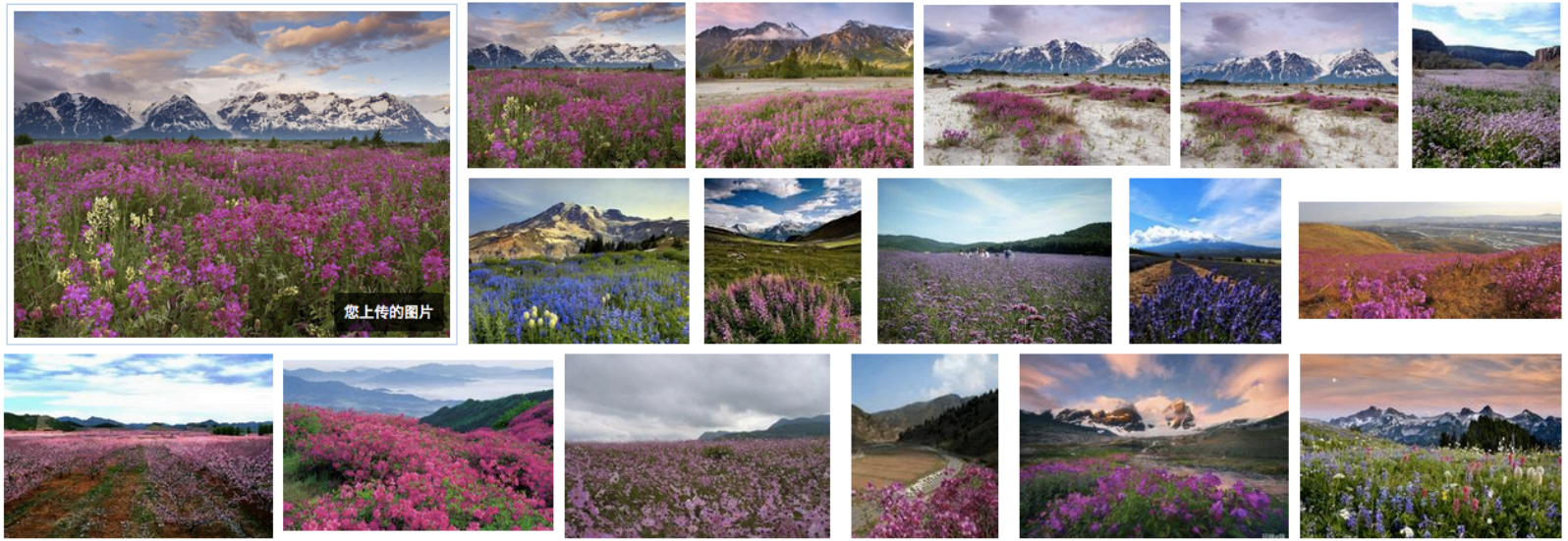
Baidu Photo Wonder

Baidu Photo Wonder

Baidu Photo Wonder

Great-China EFFIE Gold Award 2013

Visual Search

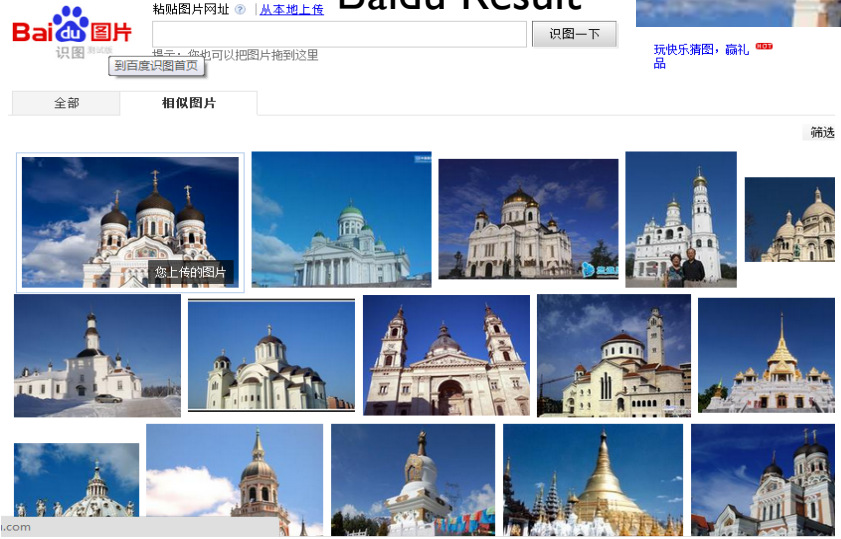


第2页

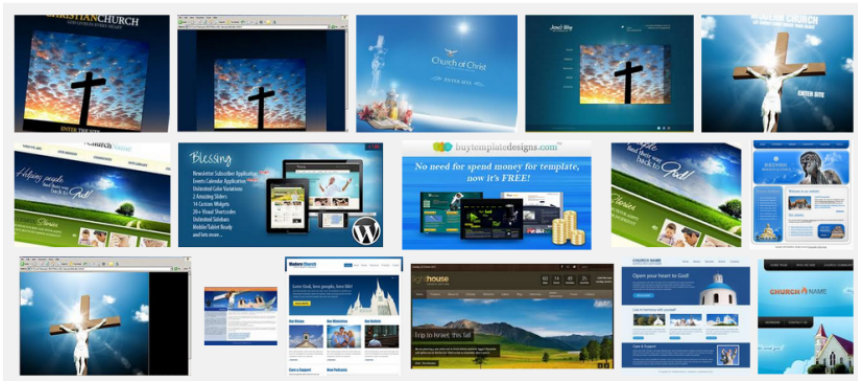
Baidu Visual Search

Query

Baidu Result



Competitor's Result



Baidu Visual Search

Query

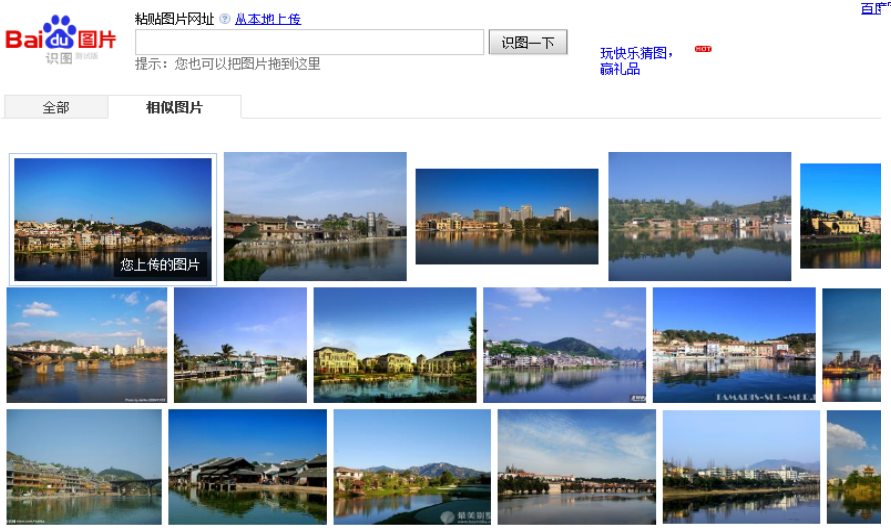
Baidu Result

Competitor's Result

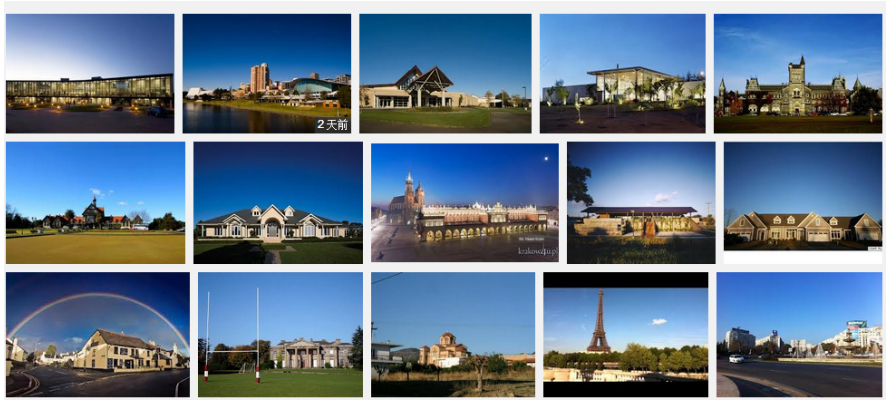
Baidu Visual Search

Query

Baidu Result



Competitor's Result



Baidu Visual Search

Qeury

Baidu Result

Competitor's Result

Poem composition based on photos

手机百度客户端 Baidu Mobile Search APP 为你写诗

诗意指数 91 分

老佛爷不听政，只愿为你写诗

相逢综艺屏， 自是多情梦。

只为心中月， 唯求梦里风。

手机百度客户端 Baidu Mobile Search APP 为你写诗

诗意指数 75 分

五花肉千金妞，不及汪伦为你写诗

富翁成乞丐， 贫穷是王侯。

世界无穷尽， 人间任自由。

手机百度客户端 Baidu Mobile Search APP 为你写诗

诗意指数 65 分

范进是为了给你写诗才疯的

同学聚会兴浓稠， 相识相逢相逢秋。

无情风雨人间事， 有意云烟万里愁。

手机百度客户端
Baidu Mobile Search APP
为你写诗

诗意指数 72 分

额真滴错咧，湘玉给你写诗咧

碧树兰花蕙芝藏， 青山绿水兰桂香。
明日黄花浑不觉， 清茶淡酒又何妨。

扫一扫 杜甫来为你写诗！
三亿人都在用的手机百度客户端～

手机百度客户端
Baidu Mobile Search APP
为你写诗

诗意指数 66 分

燕人张飞含情脉脉为你写诗

问花开未含羞草， 天涯来犹带刺花。
春来花一狮子吼， 月下无信天游夏。

扫一扫 杜甫来为你写诗！
三亿人都在用的手机百度客户端～

手机百度客户端
Baidu Mobile Search APP
为你写诗

诗意指数 100 分

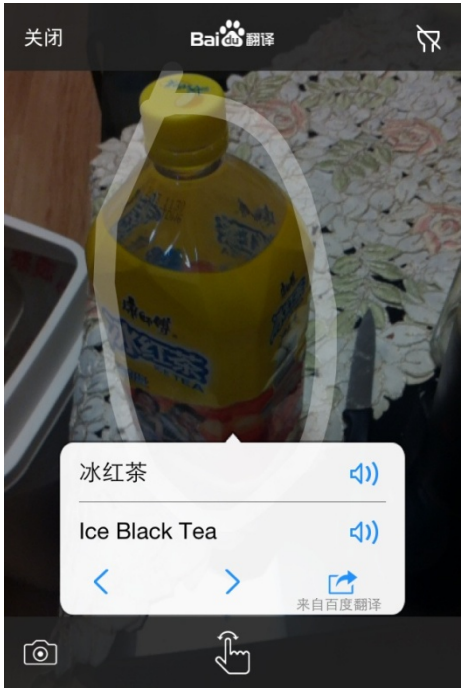
李白戒酒跪求为你写诗

莲花池畔赏荷花， 杨柳枝头映晚霞。
春雨春风春去处， 云烟过眼看梅花。

扫一扫 杜甫来为你写诗！
三亿人都在用的手机百度客户端～

Baidu Object Translation

Baidu Object Translation



Baidu Object Translation

Baidu Object Translation

Deep learning: why today?

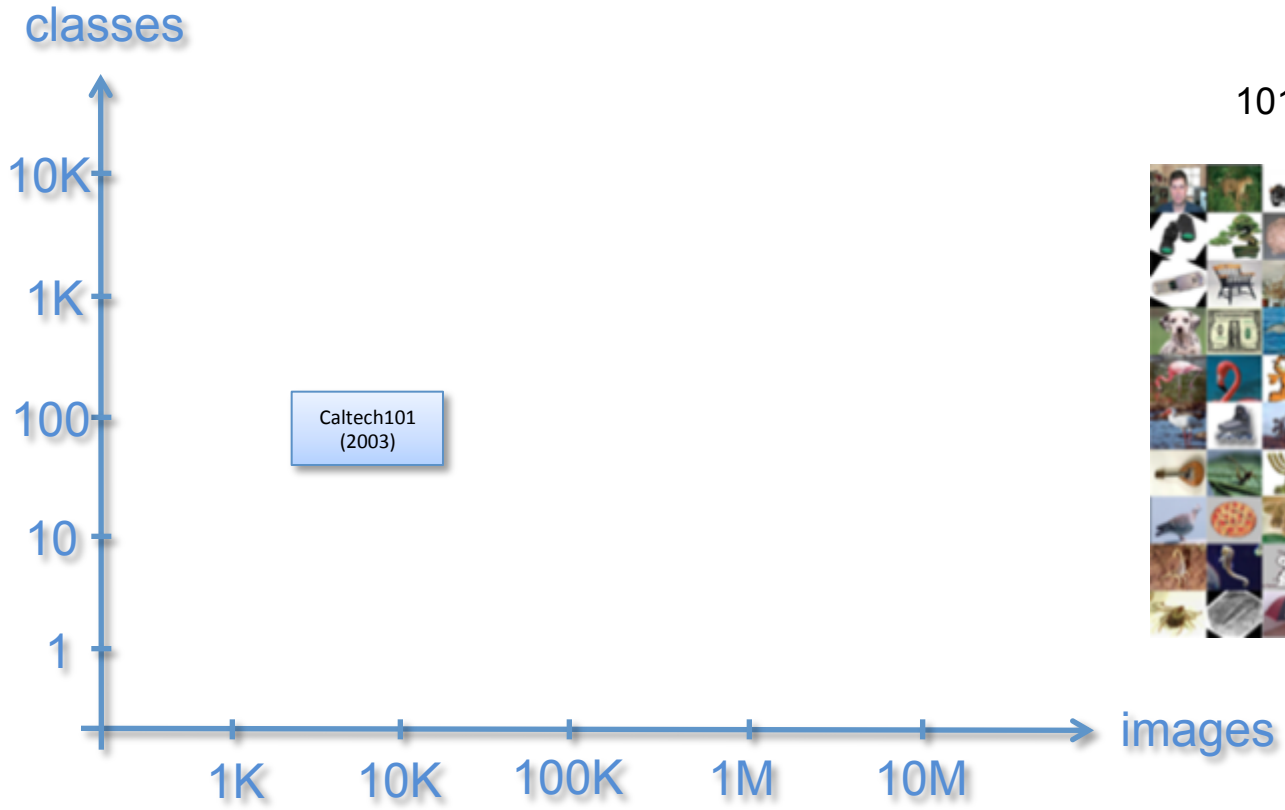
- Non-convex & non-linear
- Intensive computation
- Sensitive to initialization
- Over-fitting
- Vanishing gradient

Deep learning: why today?

- Non-convex & non-linear
- Intensive computation
- Sensitive to initialization
- Over-fitting
- Vanishing gradient

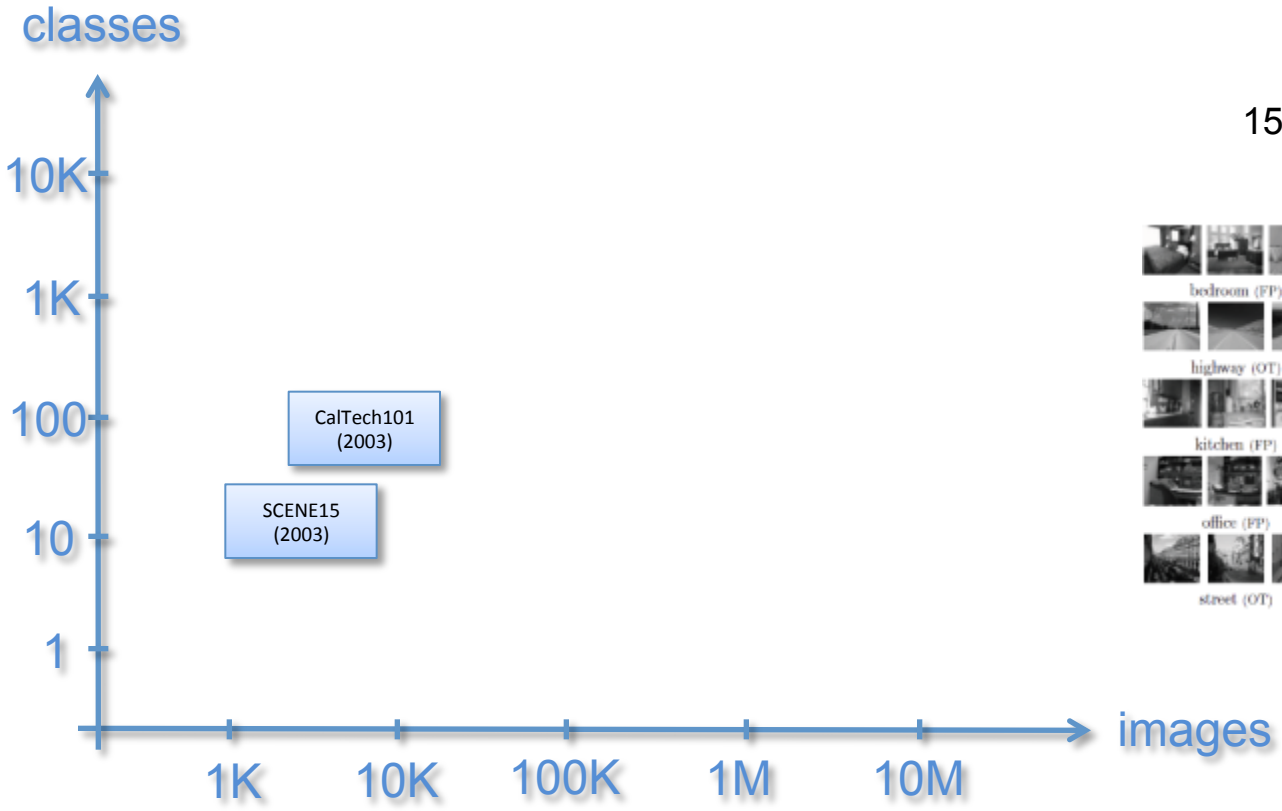
- Big data
- GPU
- Large scale parallel computation
- Layer-wise pre-training
- RELU, drop-out, better normalization, etc.

Datasets



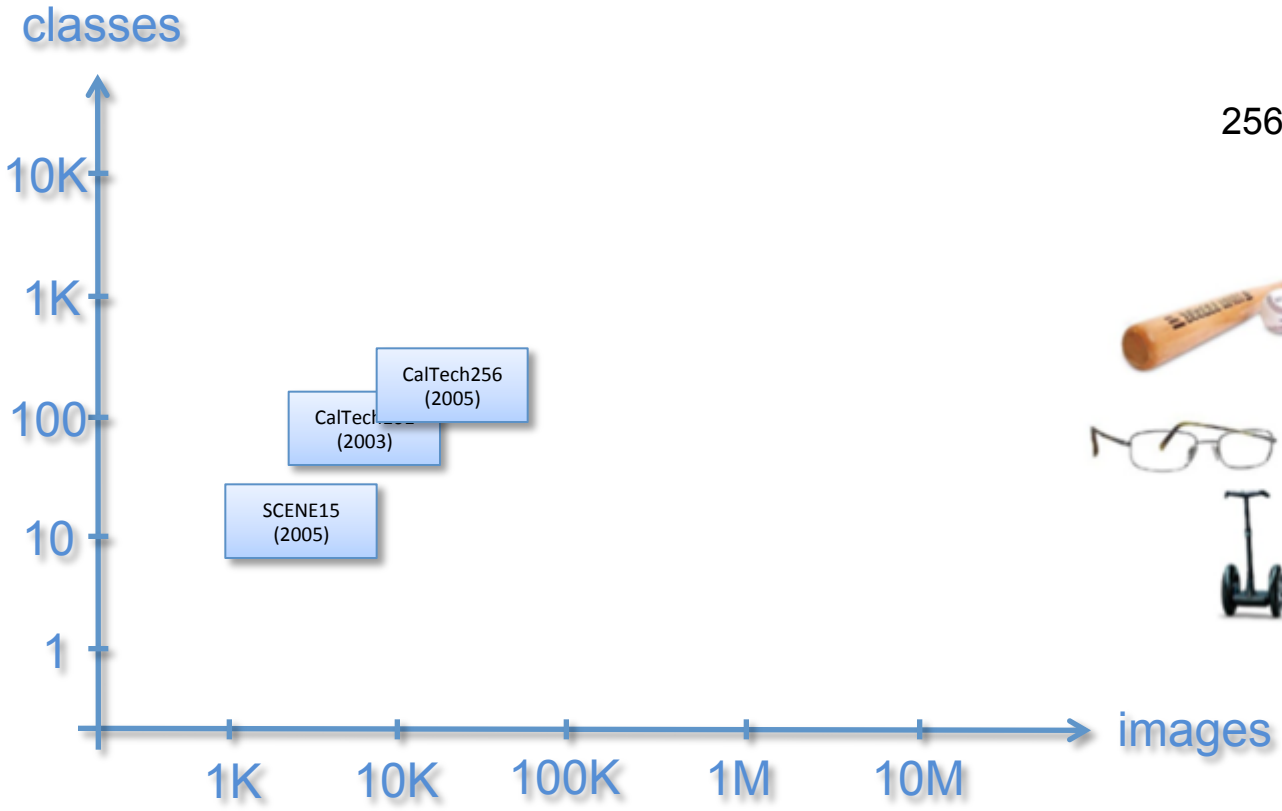
Caltech 101
101 classes, 9K images

Datasets



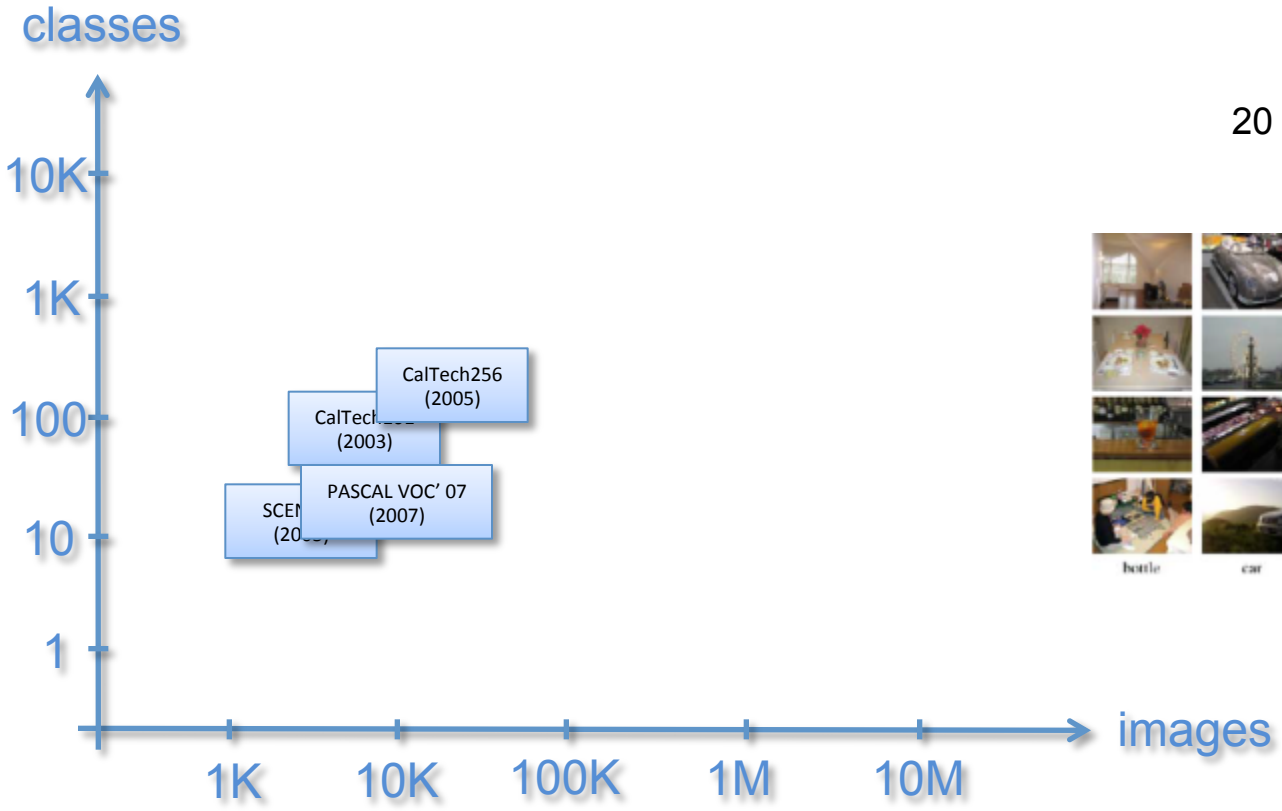
SCENE15
15 classes, 5K images

Datasets

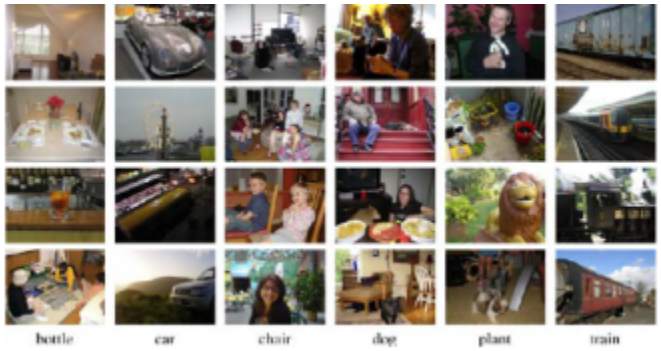


Caltech256
256 classes, 30K images

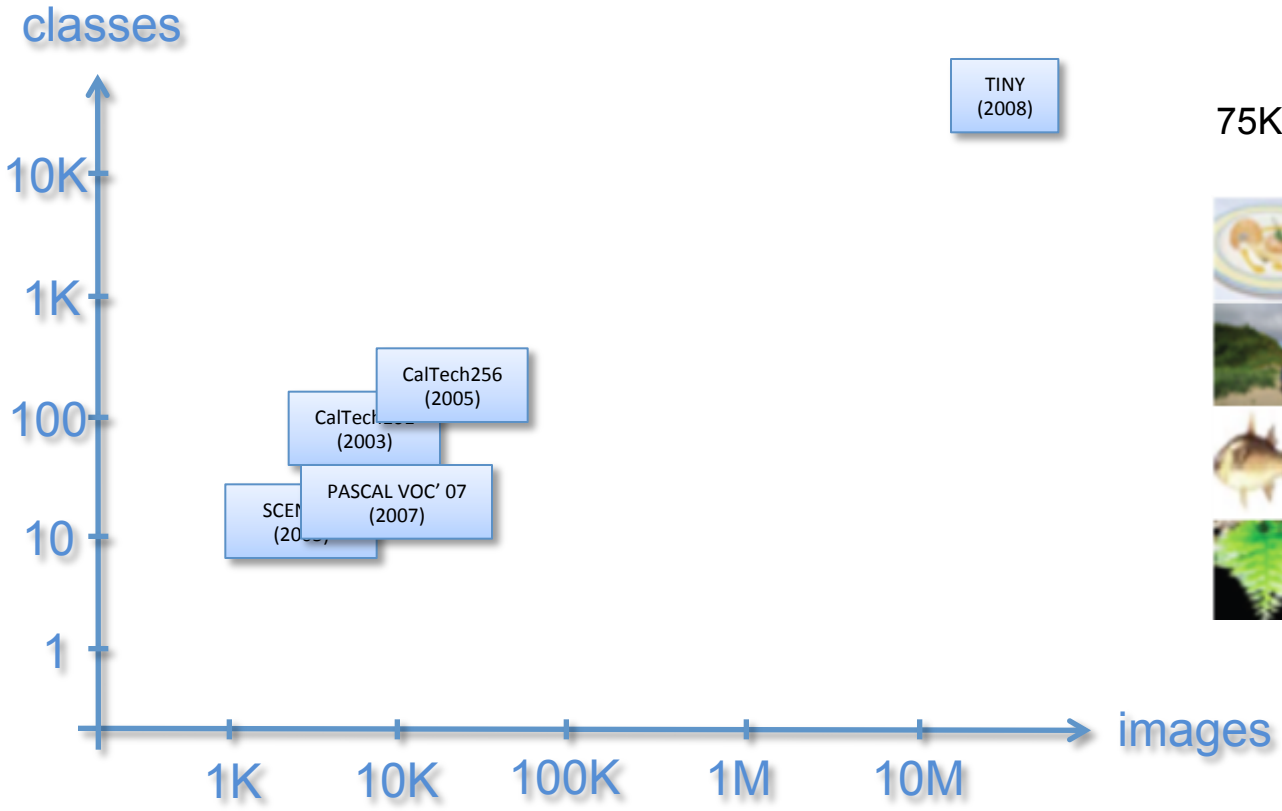
Datasets



PASCAL VOC'07
20 classes, 10K images

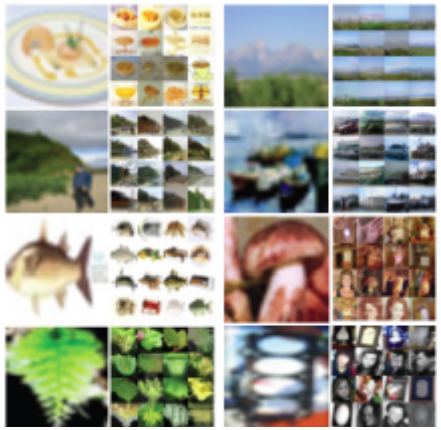


Datasets

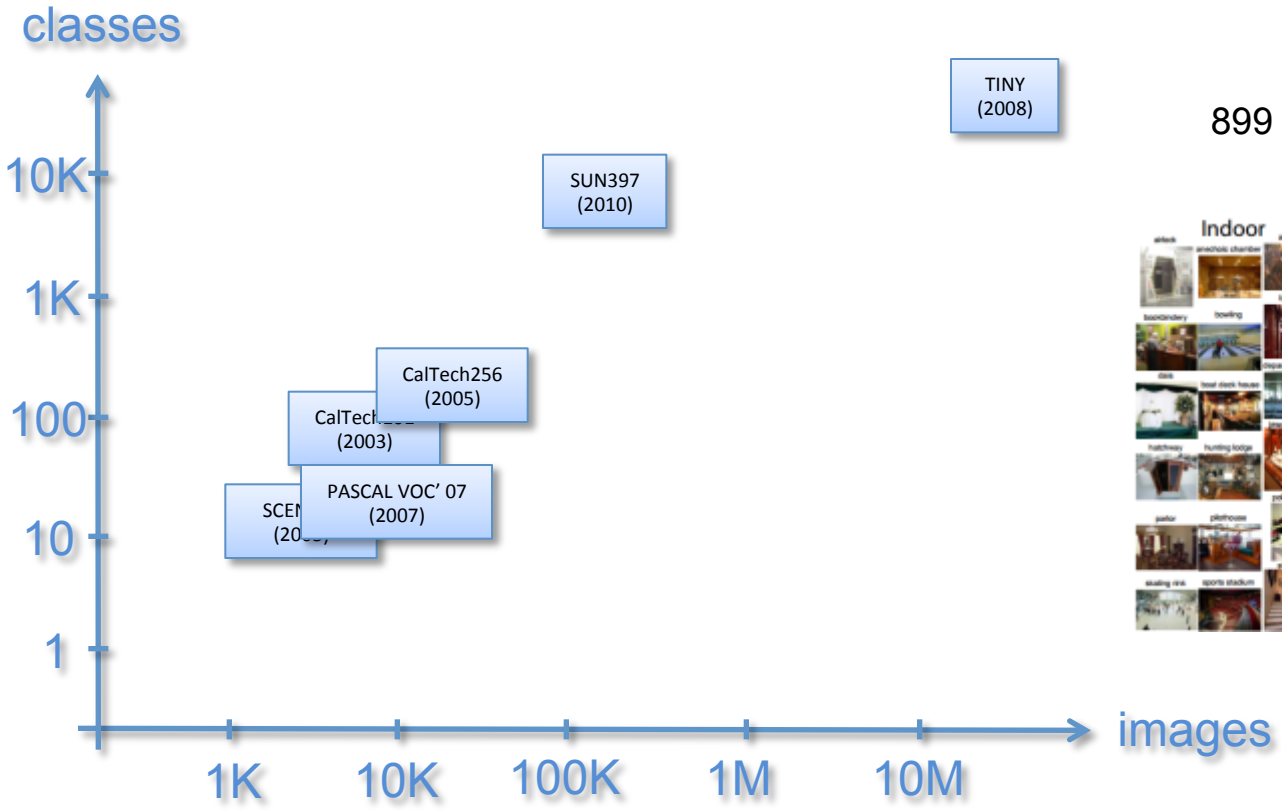


TINY
(2008)

TINY (32x32)
75K classes, 80M images

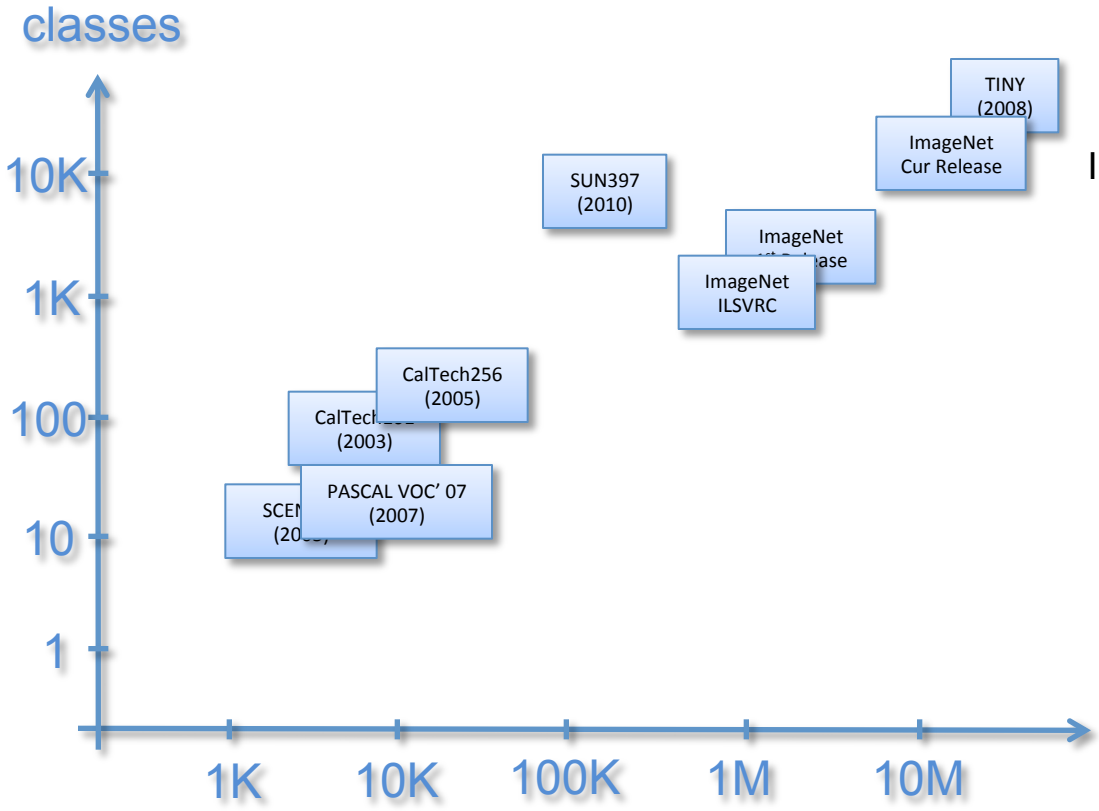


Datasets



SUN397
899 classes, 131K images

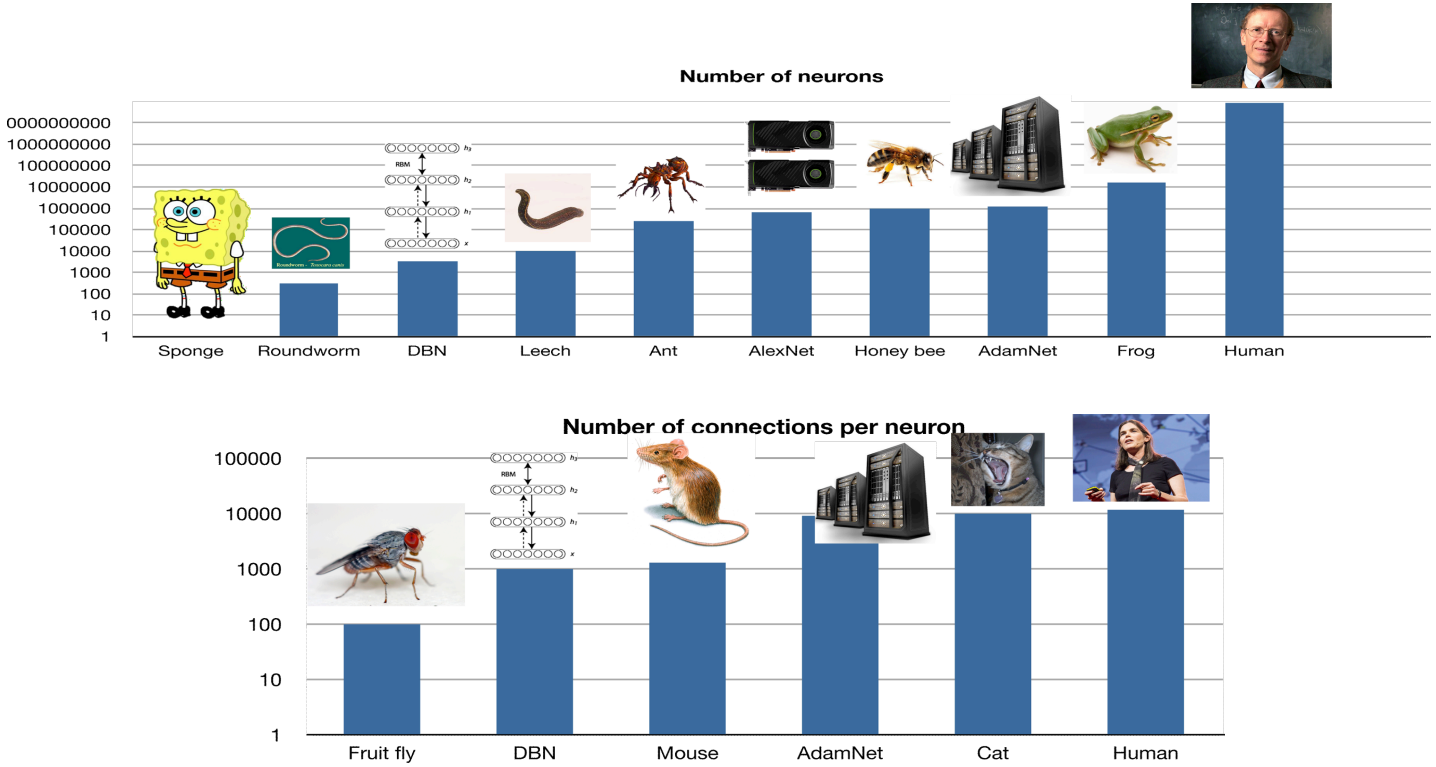
Datasets



ImageNet

1st release: 5K classes, 3M images
Cur release: 22K classes, 14M images
ILSVRC 2010: 1K classes, 1.4M images

More neurons vs. more connections

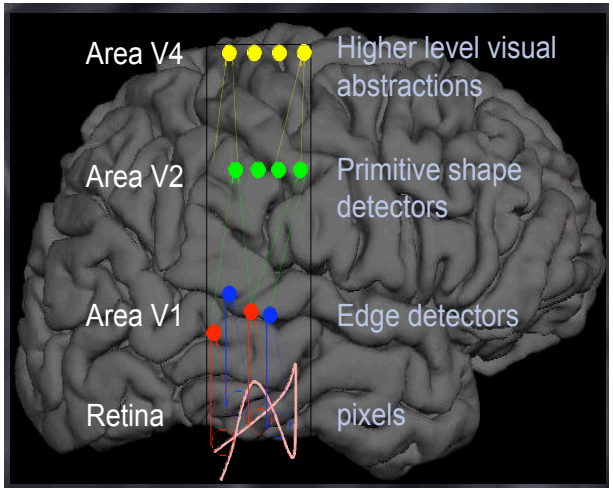


Slide credit: Ian Goodfellow

Human Brain

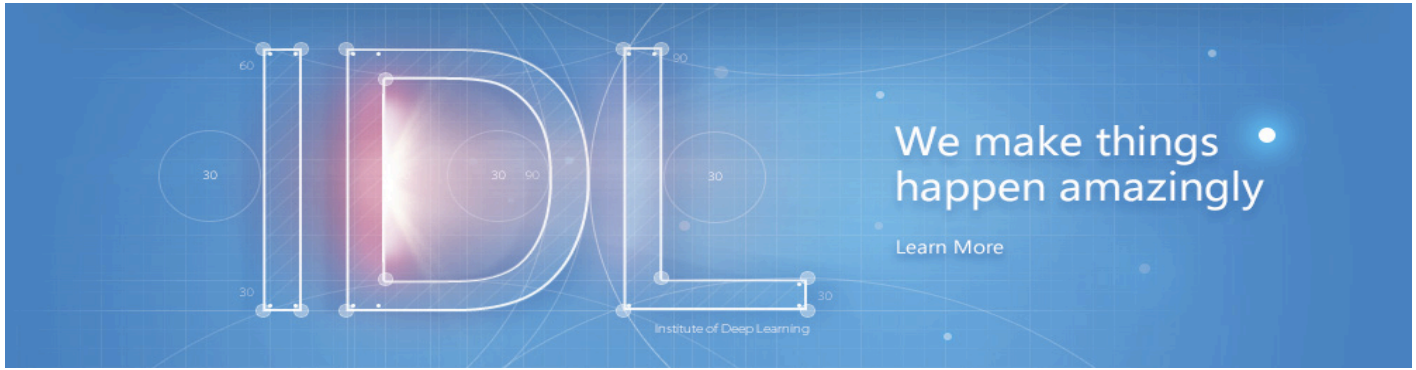
- 1.5kg, 2% of body weight, but consume 20% energy
- 100 billion neurons
- Each neuron has 5000 synapses
- firing rate: 200 per second
- Computation capacity:
  - 10^11*5000*200=10^17=100 petaflops
  - 20W, 5petaflops/w

- The most powerful supercomputer(天河二号)
  - 33.86 petaflops
  - 18*10^6w, 2.14Gflops/w



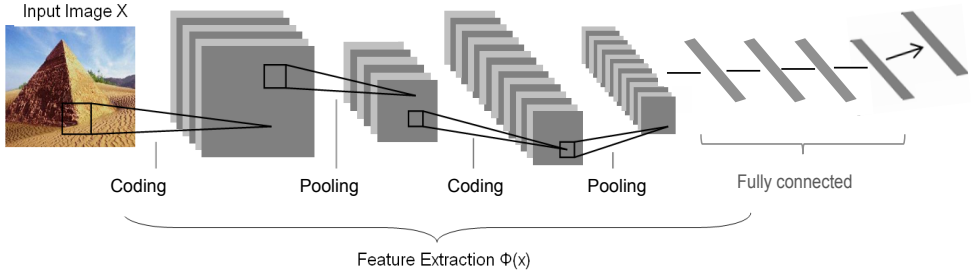
We continue to progress on

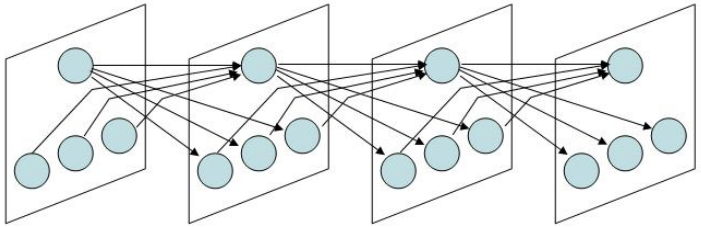
- Large-scale parallel training
- Modeling structured,unstructured, multimodality data
- New computing hardware for deep learning
- High-performance computing
- Neural science, …

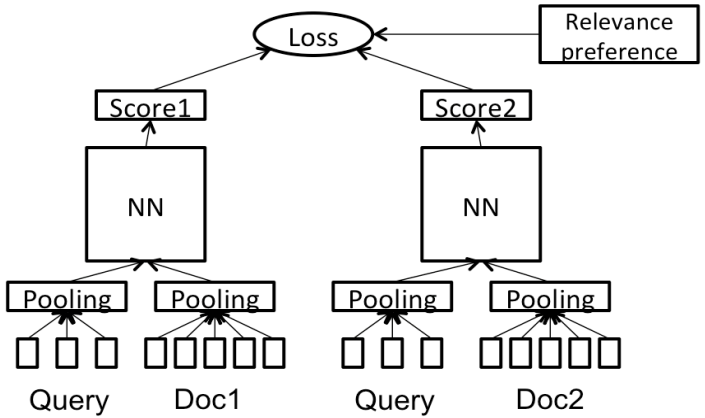


Clarification: Deep Learning is NOT Blackbox

- DL is a language, just like graphical models
- Prior knowledge: in model structure, not feature engineering

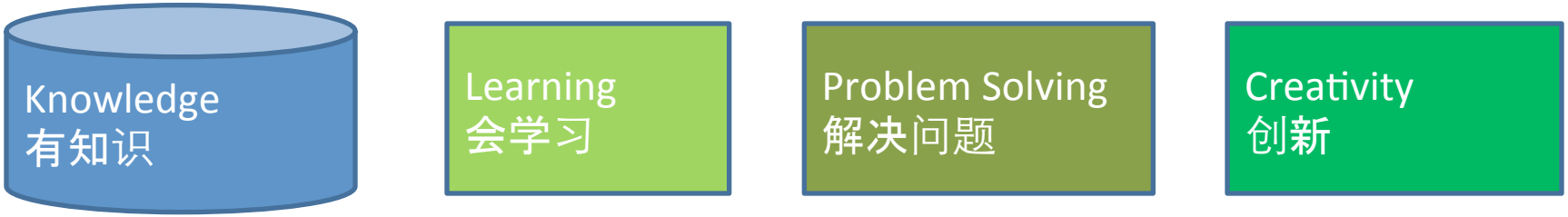




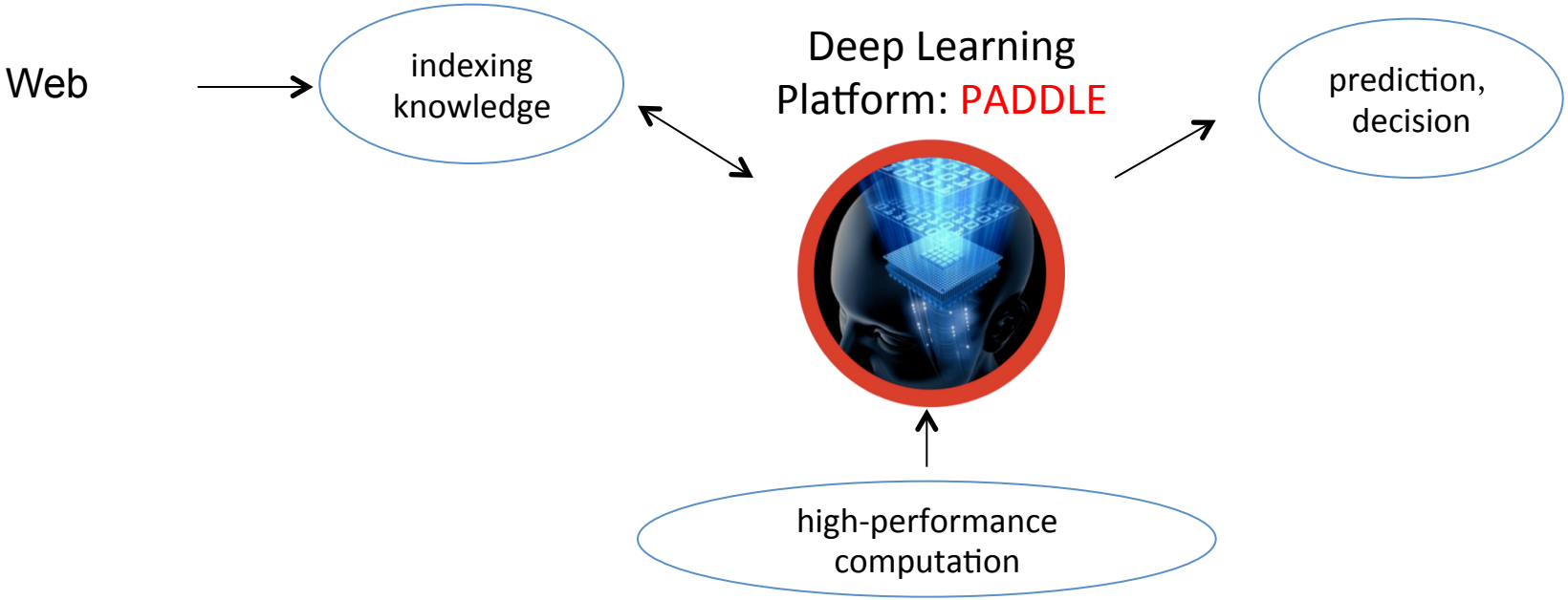


Clarification: Deep Learning is NOT AI

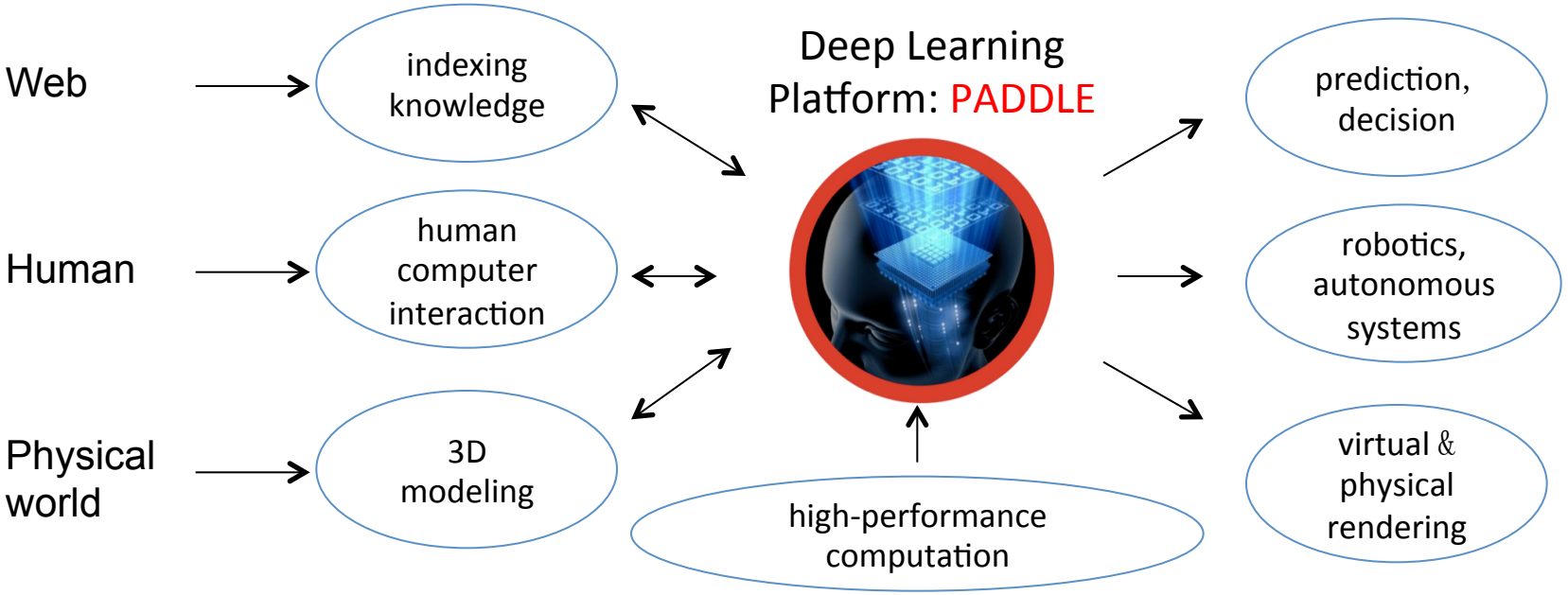
- Deep learning might be our current best shot towards AI
- AI is our ultimate goal
- What's inside AI?



Technology portfolio at Baidu IDL



Technology portfolio at Baidu IDL



Summary

- Deep Learning made big success at Baidu
- New paradigm of AI: big data + complex models
- Computation capacity enables many things to happen

We are hiring (Beijing & Silicon Valley)…

- Machine Learning
- Big Data Analytics
- Human-Computer Interaction
- Robotics
- Computer Vision
- 3D Vision

idl_job@baidu.com


Bai
du
百度