Deep Neural Networks:

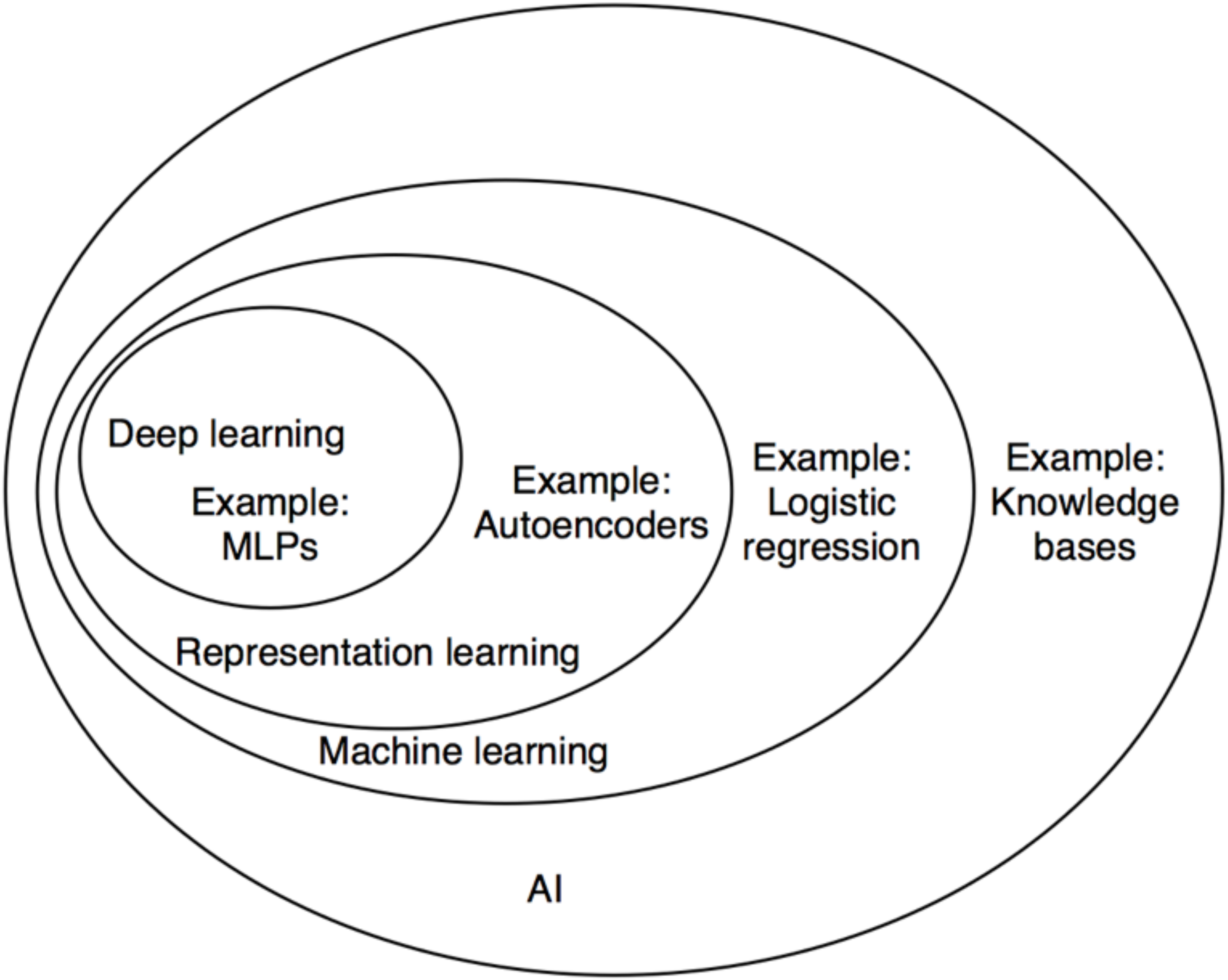
Part I What are DNN and Deep Learning

Yuan-Kai Wang, 2016

Web site of this course: http://pattern-recognition.weebly.com

source: Deep learning: a birds-eye view, by R. Pieters, 2015.

Machine Learning vs. Deep Learning



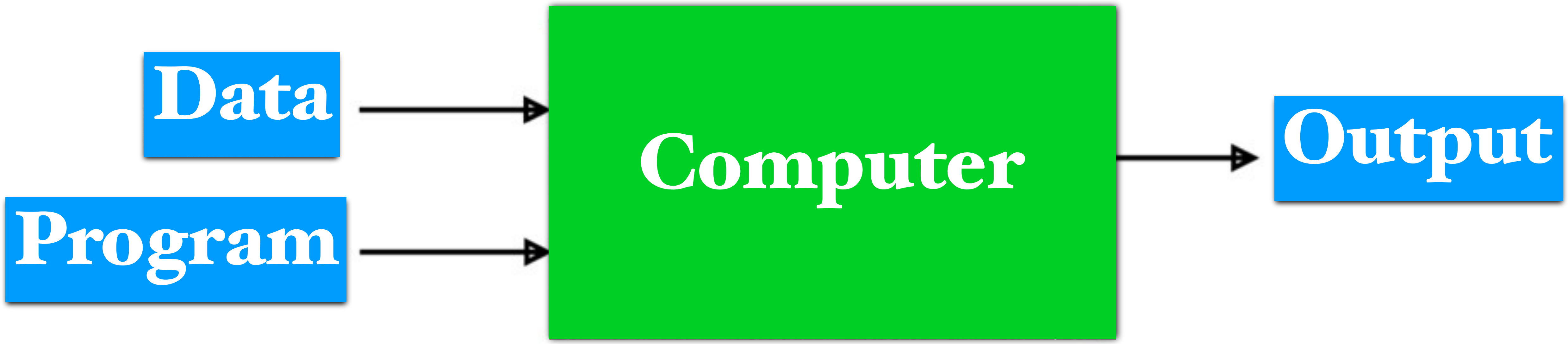
Deep Learning vs. Representation Learning

Representation learning
Attempts to automatically learn good features or representations

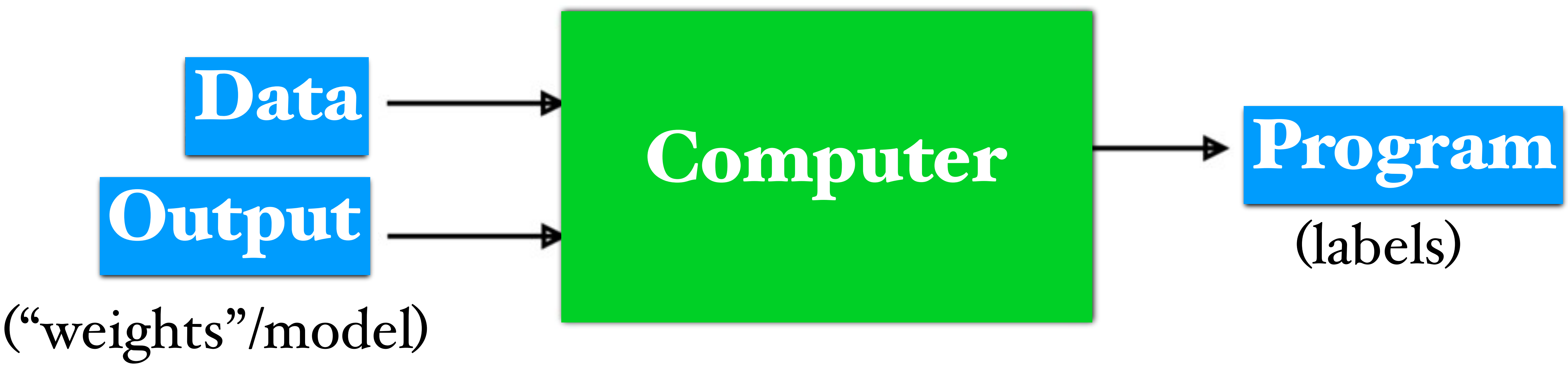
Deep learning
Attempt to learn multiple levels of representation of increasing complexity/abstraction

Traditional Programming vs. Machine Learning

Traditional Programming:

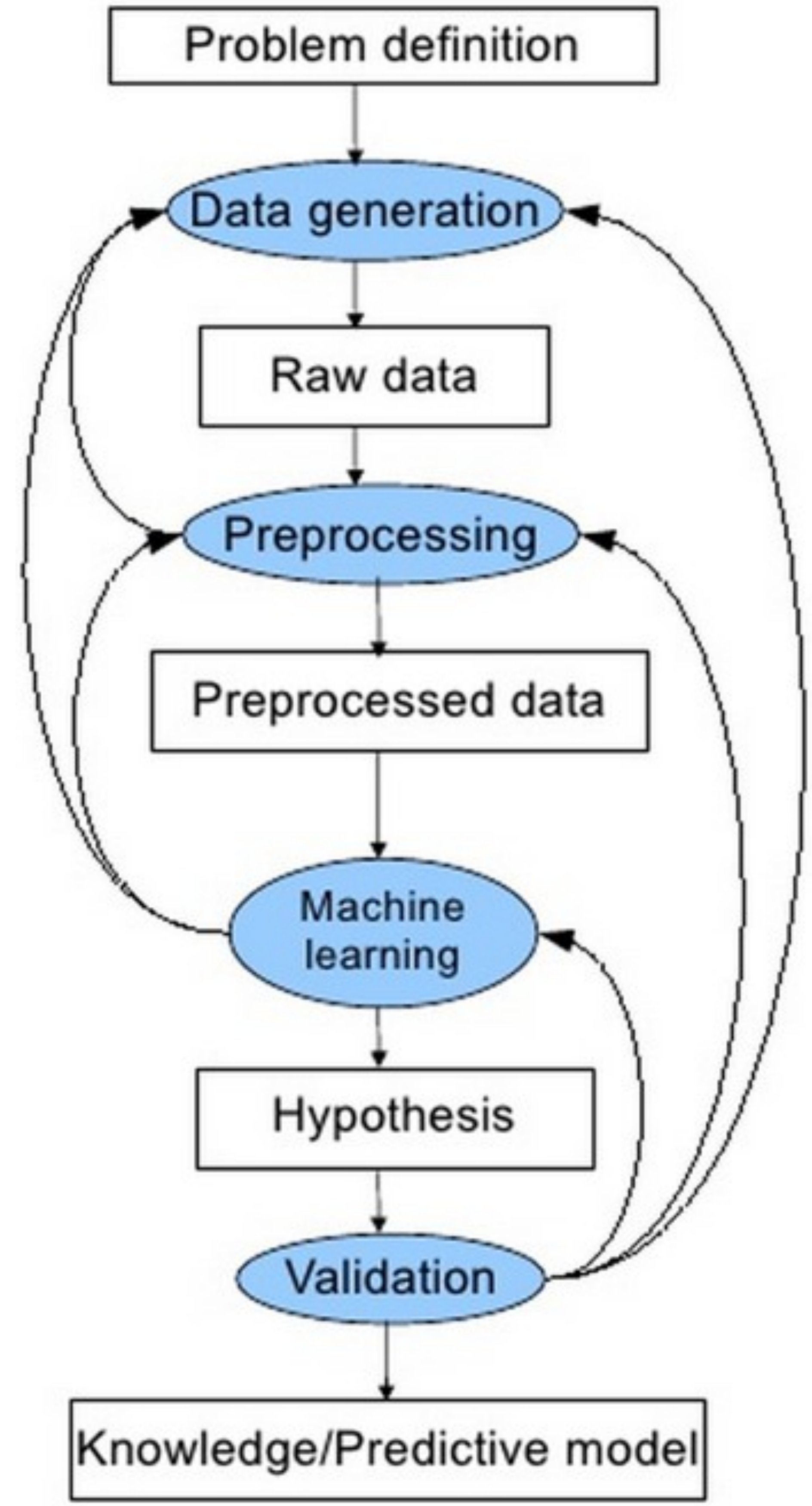


Machine Learning:



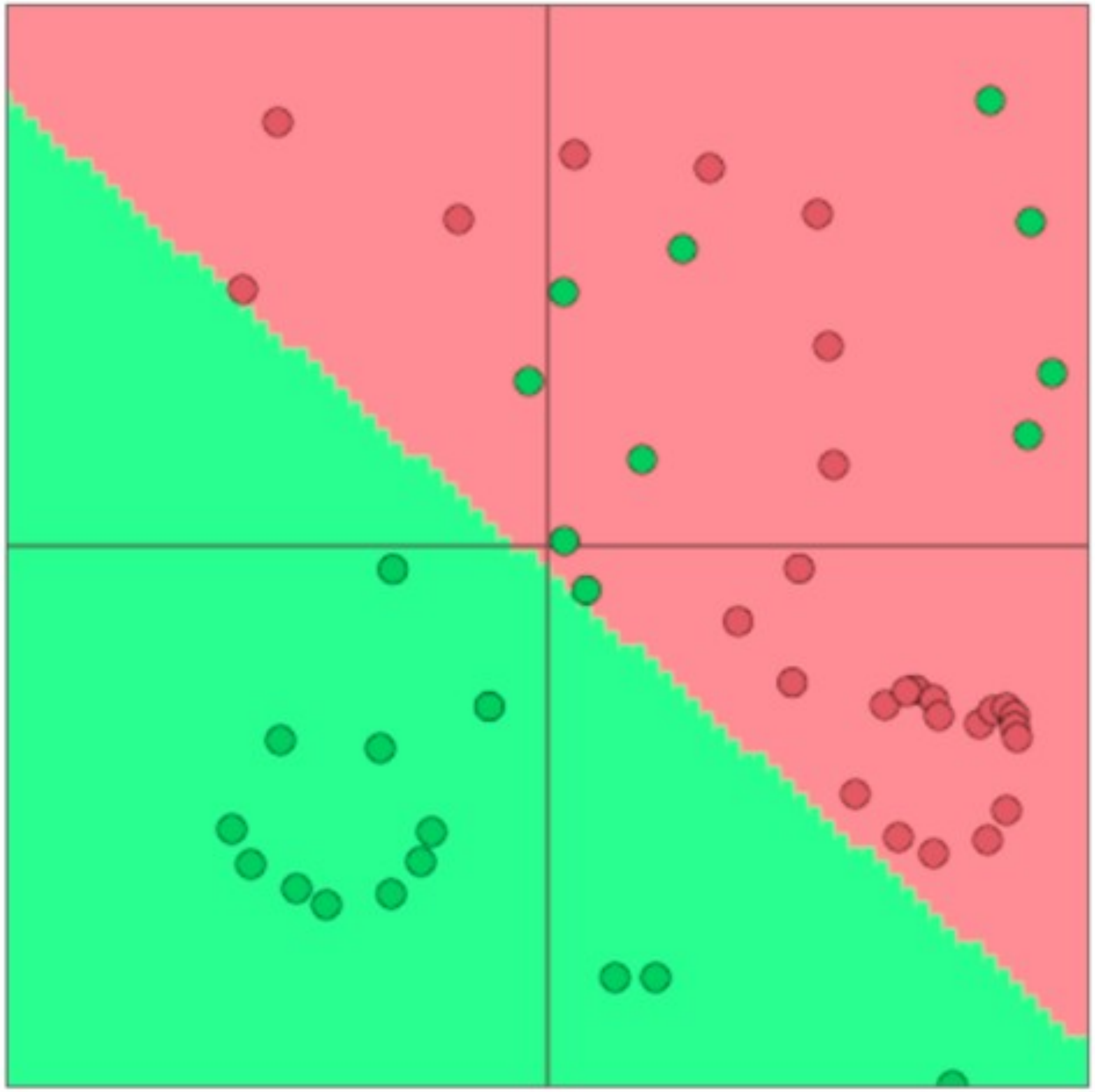
Machine Learning

- Most machine learning methods work well because of human-designed/hand-engineered features (representations)
- machine learning -> optimising weights to best make a final prediction

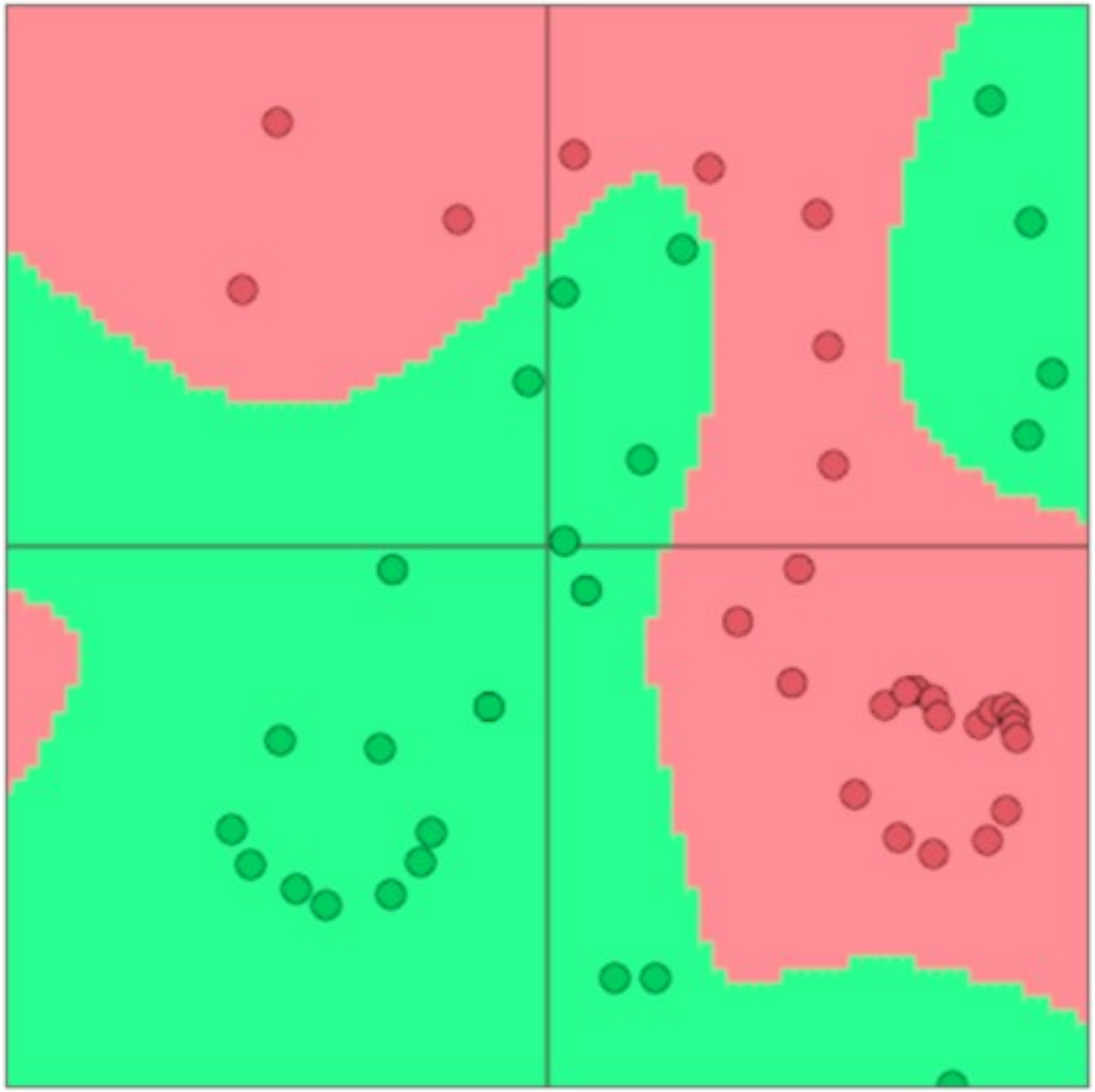


Deep Learning: Why?

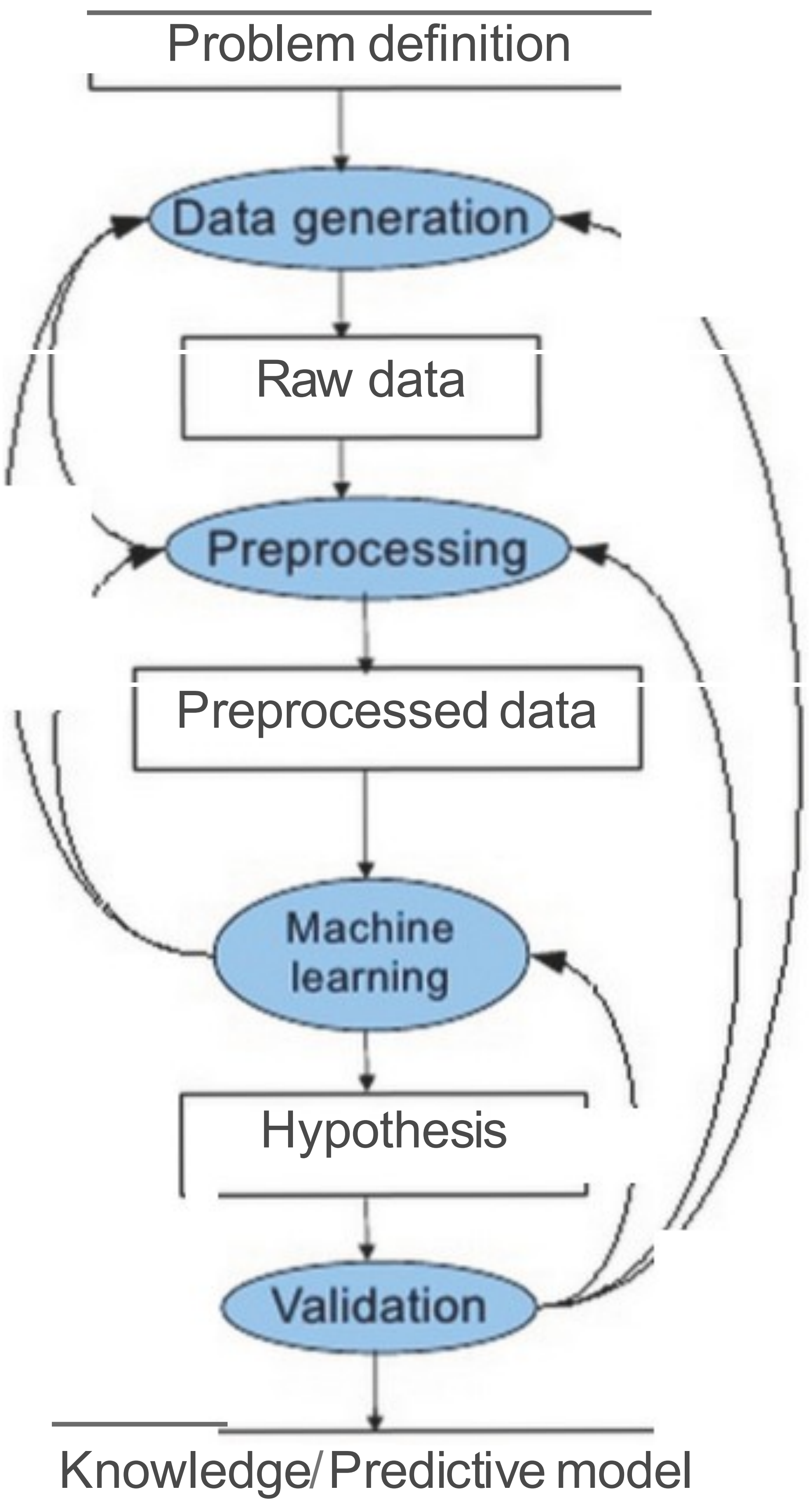
Typical ML Regression



Neural Net



Machine Learning → Deep Learning



One pa此 of the data mlnlng process

- Each step generates many questions:

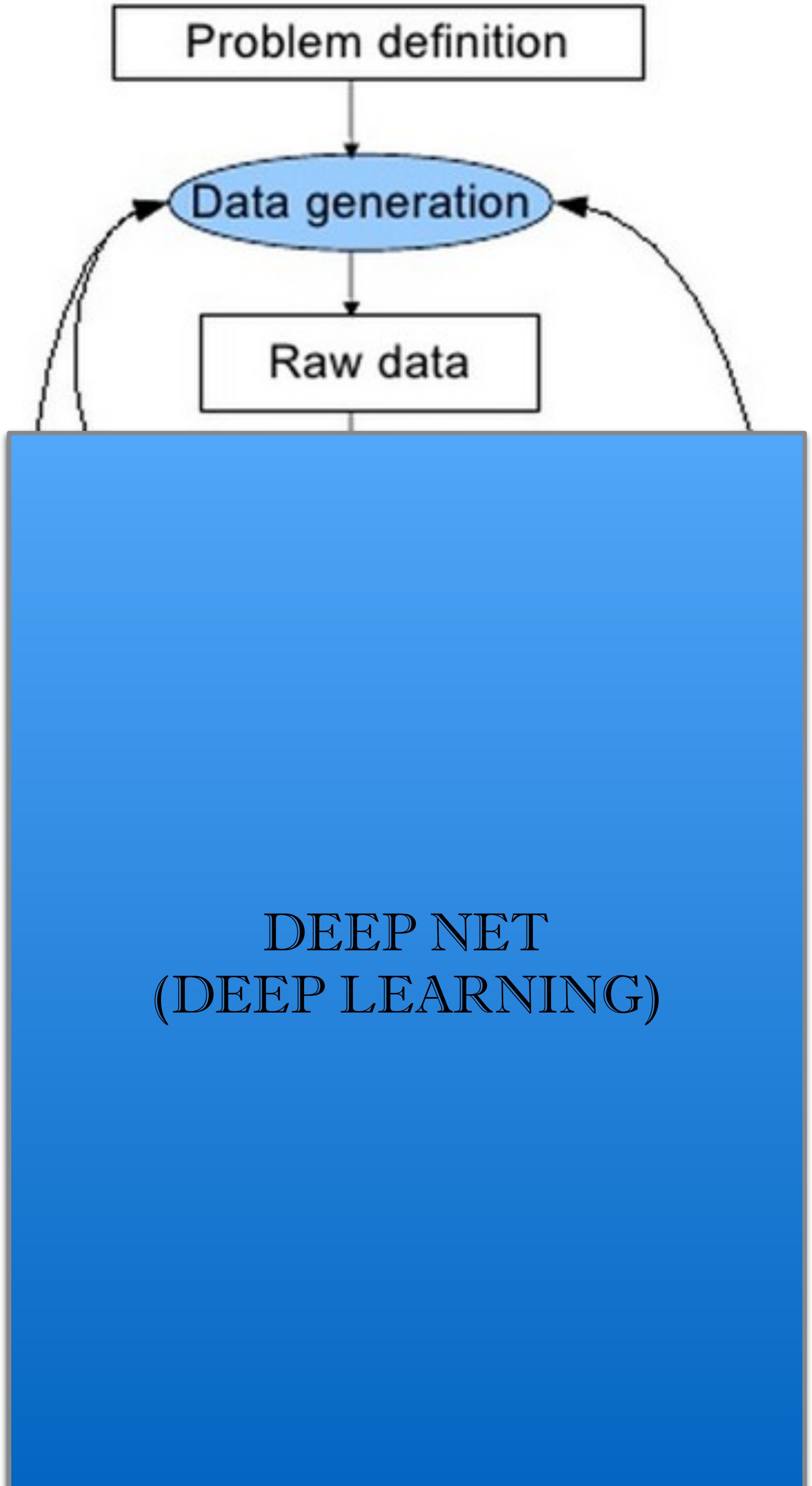
  Data generation: data types, sample size, online/offline...

  Preprocessing: normalization, missing values, feature selection/extraction...

  Machine learning: hypothesis, choice of learning paradigm/algorithm ...

  Hypothesis validation: cross-validation, model deployment...

Machine Learning → Deep Learning

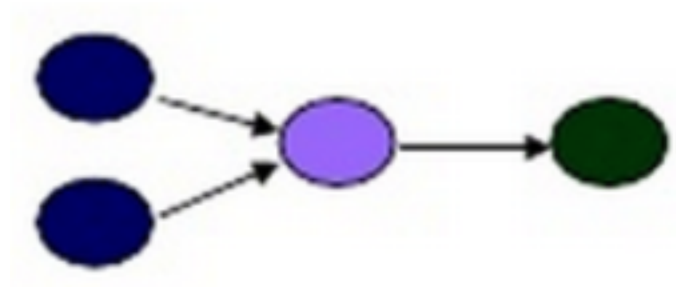


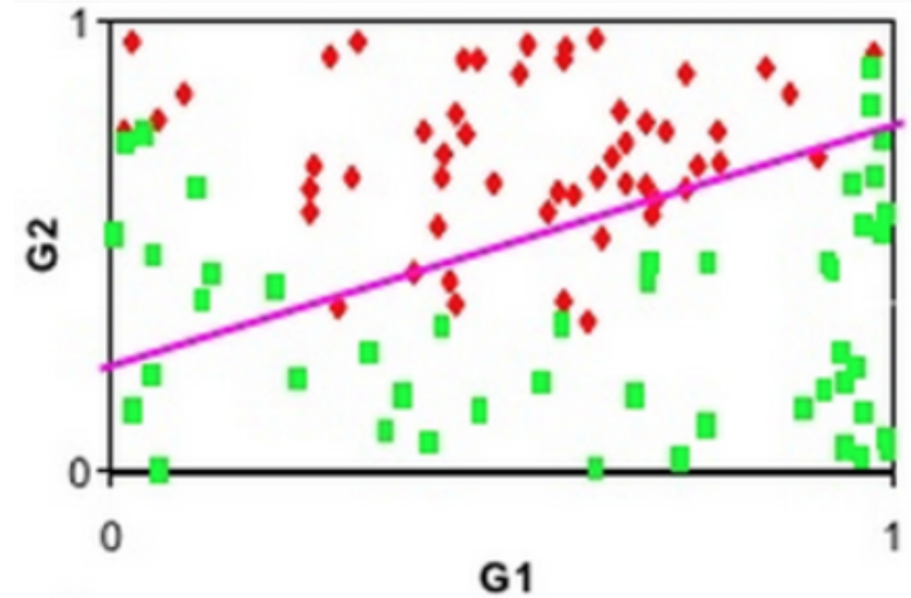
One part of the data mining process

- Each step generates many questions:
  - Data generation: data types, sample size, online/offline...
  - Preprocessing: normalization, missing values, feature selection/extraction...
  - Machine learning: hypothesis, choice of learning paradigm/algorithm...
  - Hypothesis validation: cross-validation, model deployment...

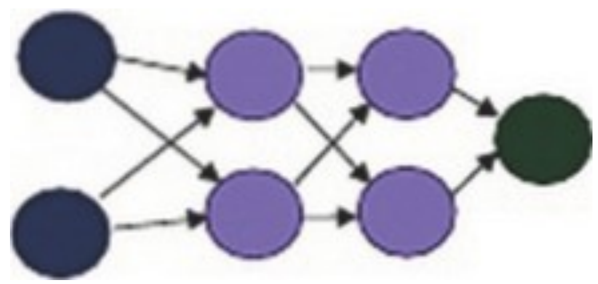
Deep Learning: Why?

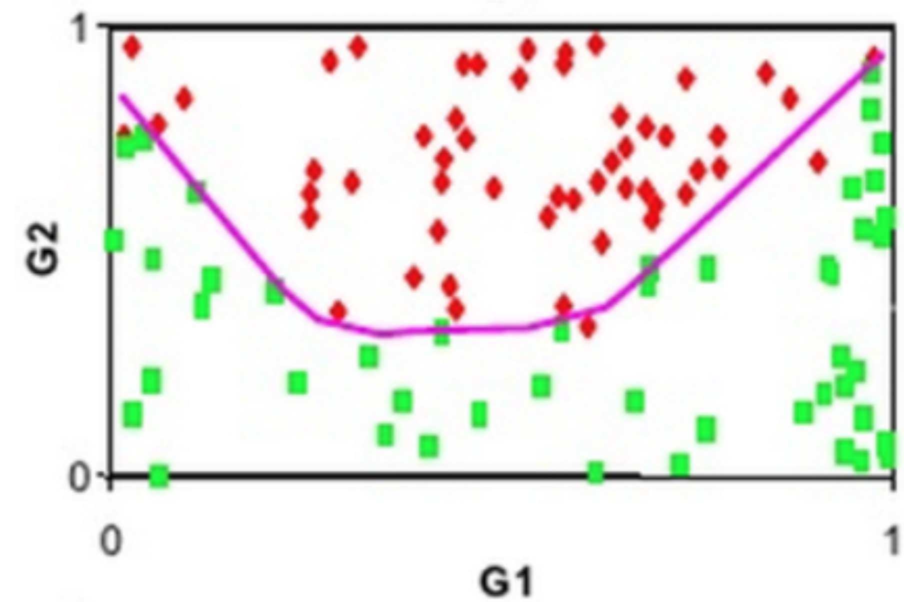
1 neuron



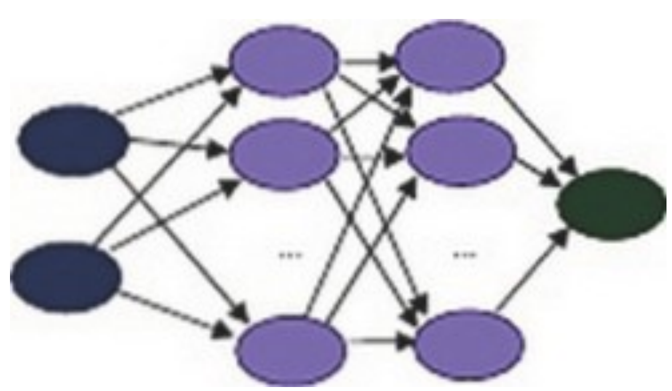


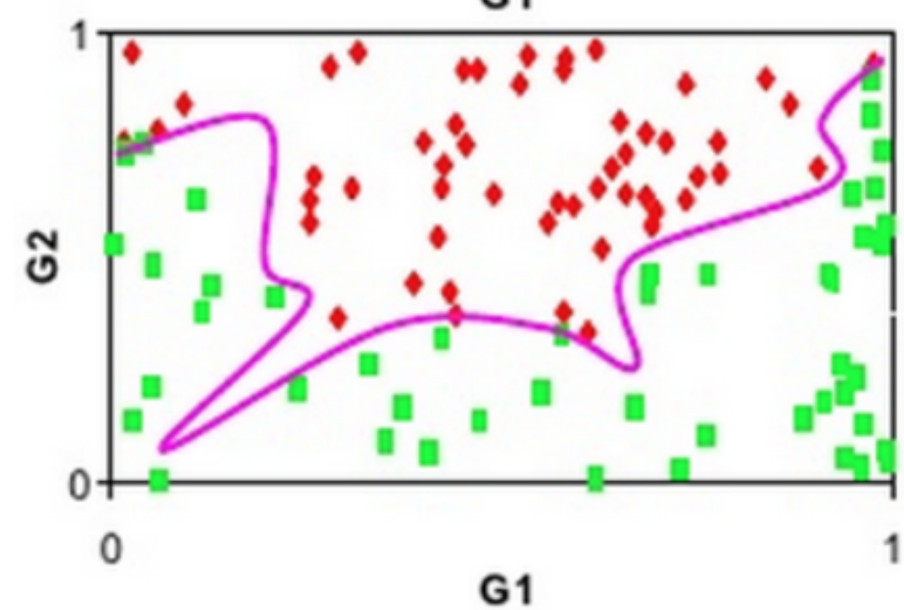
2 + 2 neurons



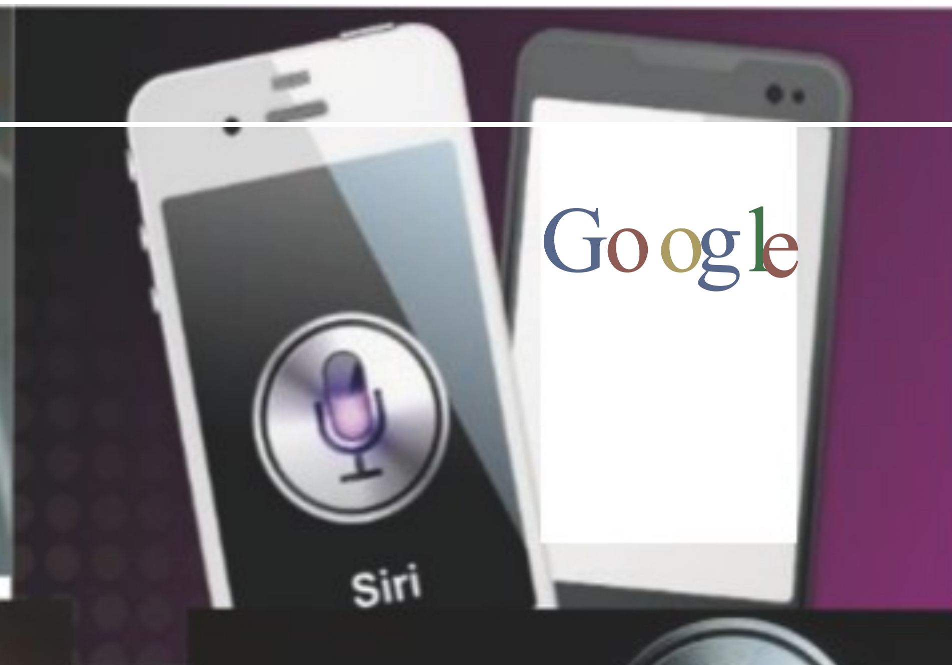


10 +10 neurons

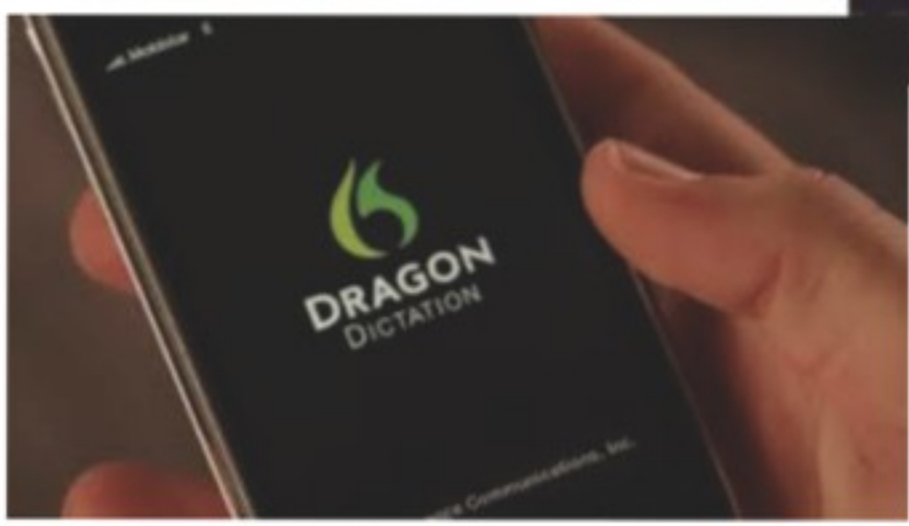




Deep Learning is everywhere...


Google
Siri


Google now


DRAGON
DICTATION

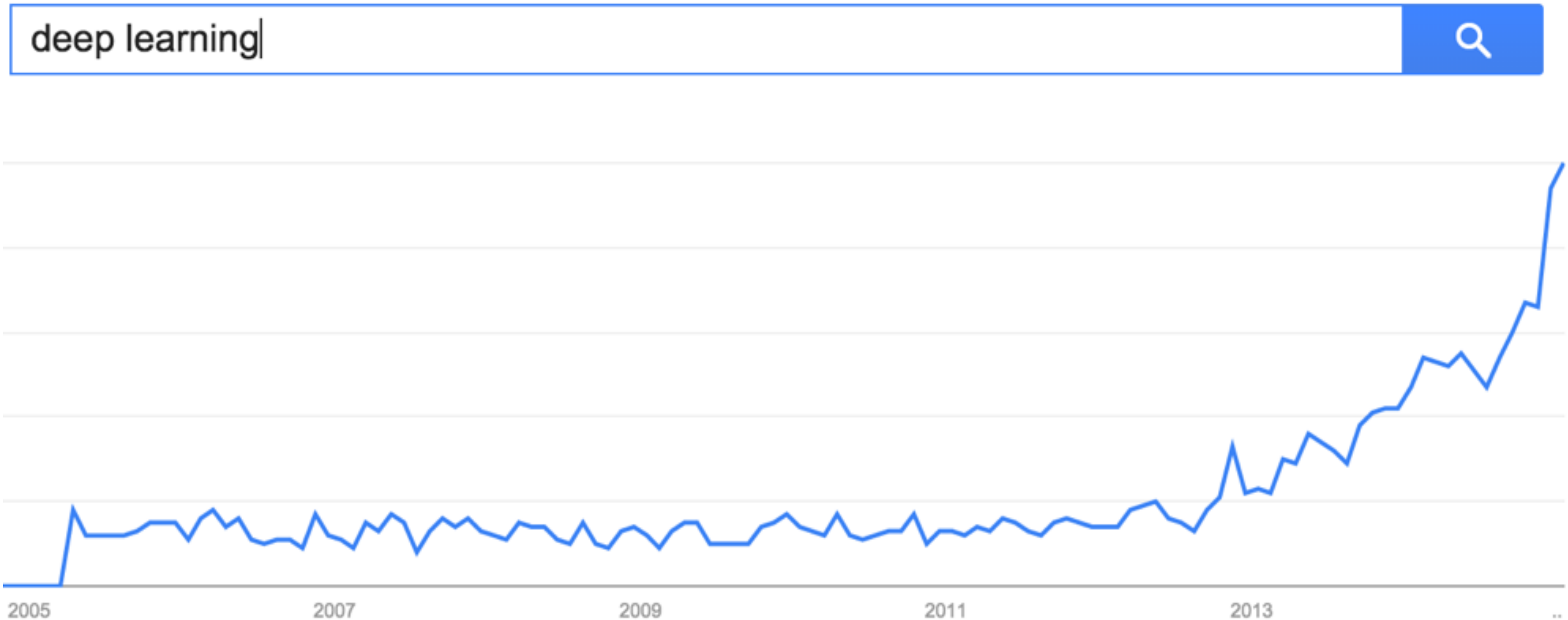

Siri


中国联通 18:15
讯飞语点
您需要做什么?
打开手机中国
正在打开手机中国
点击说话


XBOX ! BRINC
ME A PIE!

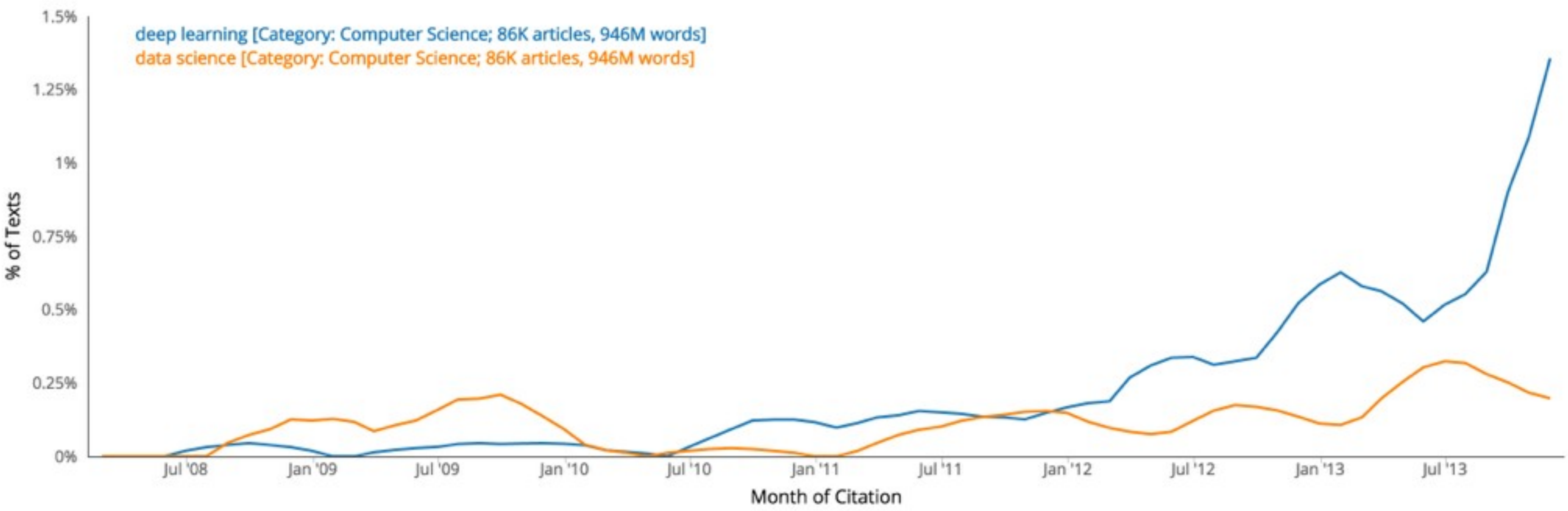

Bai du 百度
CARRIER
12:00 PM

Deep Learning in the News



(*source: Google Trends*)

Scientific Articles:



(source: arXiv bookworm, Culturomics)

NYU "Deep Learning" Professor LeCun Will Head Facebook's New Artificial Intelligence Lab

Posted Dec 9, 2013 by **Josh Constine** (@joshconstine)

By teaching a computer to think, Facebook hopes to better understand how its users do too. So today the company announced that one of the world's leading deep learning and machine learning scientists, NYU's Professor Yann LeCun, will lead its new artificial intelligence laboratory.

MIT Technology Review first reported that Facebook would launch an Artificial Intelligence lab back in September, but now it has something of a celebrity scientist at its helm. Facebook's AI research will be split across its Menlo Park headquarters, London office, and a new AI lab built just a block from NYU's campus in Manhattan.

LeCun has been pioneering artificial intelligence breakthroughs since the 1980s when he developed an early version of the "back-propagation algorithm" that became the top way to train artificial neural networks. He went on to work for AT&T Bell Laboratories where he created the "convolutional network model" that mimics the visual cortex of living beings to create a pattern recognition system for machines. This model was used for optical character recognition and handwriting recognition that powered how many banks read checks in the late 1990s and early 2000s.

LeCun's expertise is in "deep learning" speech and image recognition systems has driven his research in building visual navigation systems for self-driving cars, autonomous ground robots, drones, and more.

Big Data EXCLUSIVE VB

Snapchat is quietly building a research team to do deep learning on images, videos

April 8, 2015 3:15 PM

Jordan Novet

Snapchat, that well-funded Los Angeles startup providing a popular ephemeral-mobile-messaging app, has been slowly developing a research arm to run sophisticated algorithms on user data like images and videos, VentureBeat has learned.

Google's DeepMind wins historic Go contest 4-1

W by Matt Burgess, wired.co.uk

March 15

Lee Se-dol as he concedes the final match against the AIDeepMind
Photo by: DeepMind

Why Now?

- Inspired by the architectural depth of the brain, researchers wanted for decades to train deep multi-layer neural networks.
- No successful attempts were reported before 2006 … Exception:convolutional neural networks, LeCun 1998
- SVM: Vapnik and his co-workers developed the Support Vector Machine (1993) (shallow architecture).
- Breakthrough in 2006!

Renewed Interest: 1990s

Neural networks can

- Learn multiple layers with "back propagation"
- Learn any function "theoretically"

But...

- Neural network learning is very slow and inefficient
- Researchers turn to SVMs, random forests, ...

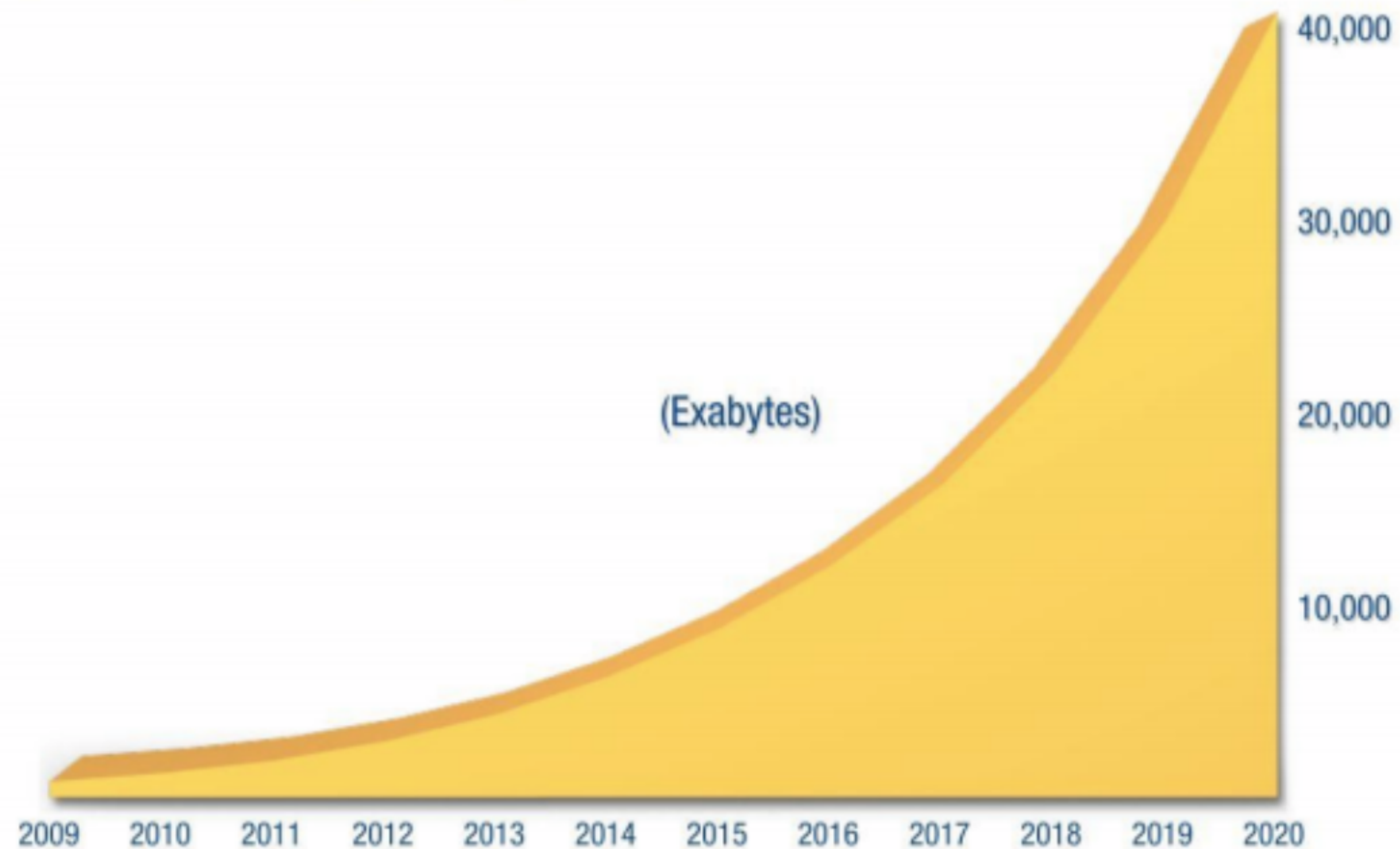
2006 Breakthrough

- More data
- Faster hardware: GPU's, multi-core CPU's
- Working ideas on how to train deep architectures

2006 Breakthrough

- **More data**
- Faster hardware: GPU's, multi-core CPU's
- Working ideas on how to train deep architectures

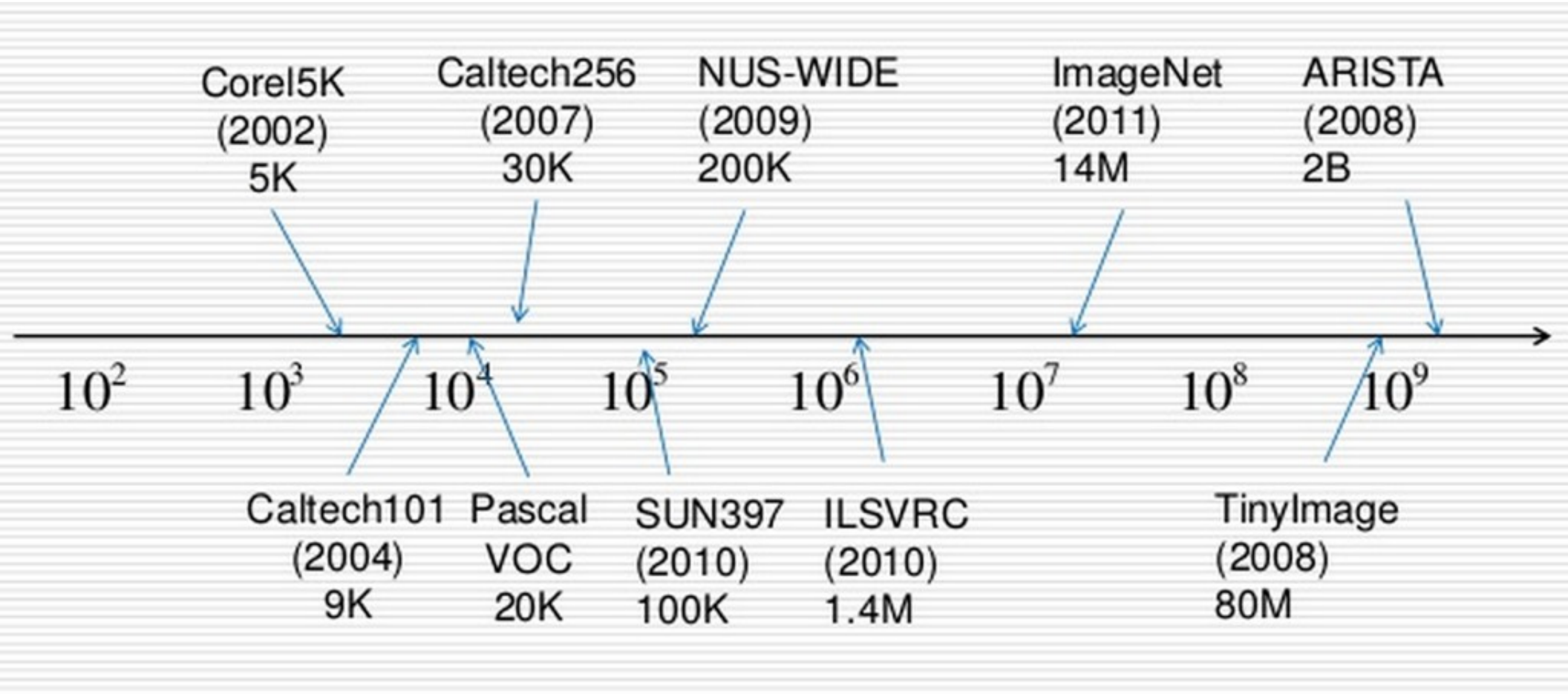
2006 Breakthrough



Source: IDC's Digital Universe Study, sponsored by EMC, December 2012

2006 Breakthrough

Growth of datasets

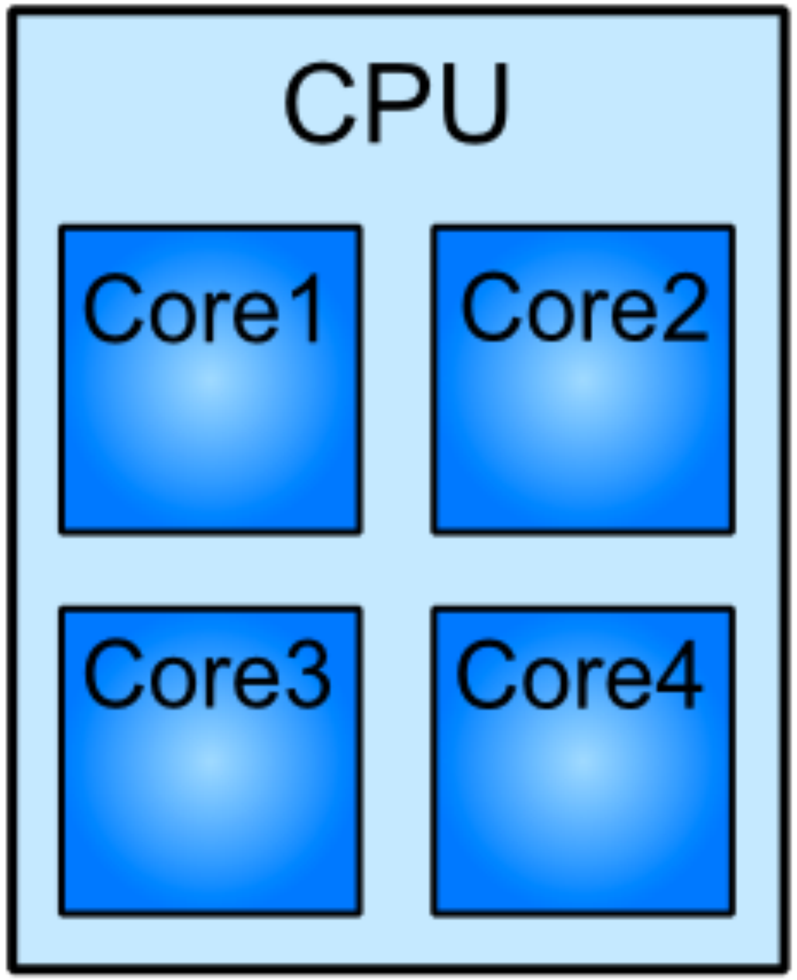


The Big Data Era

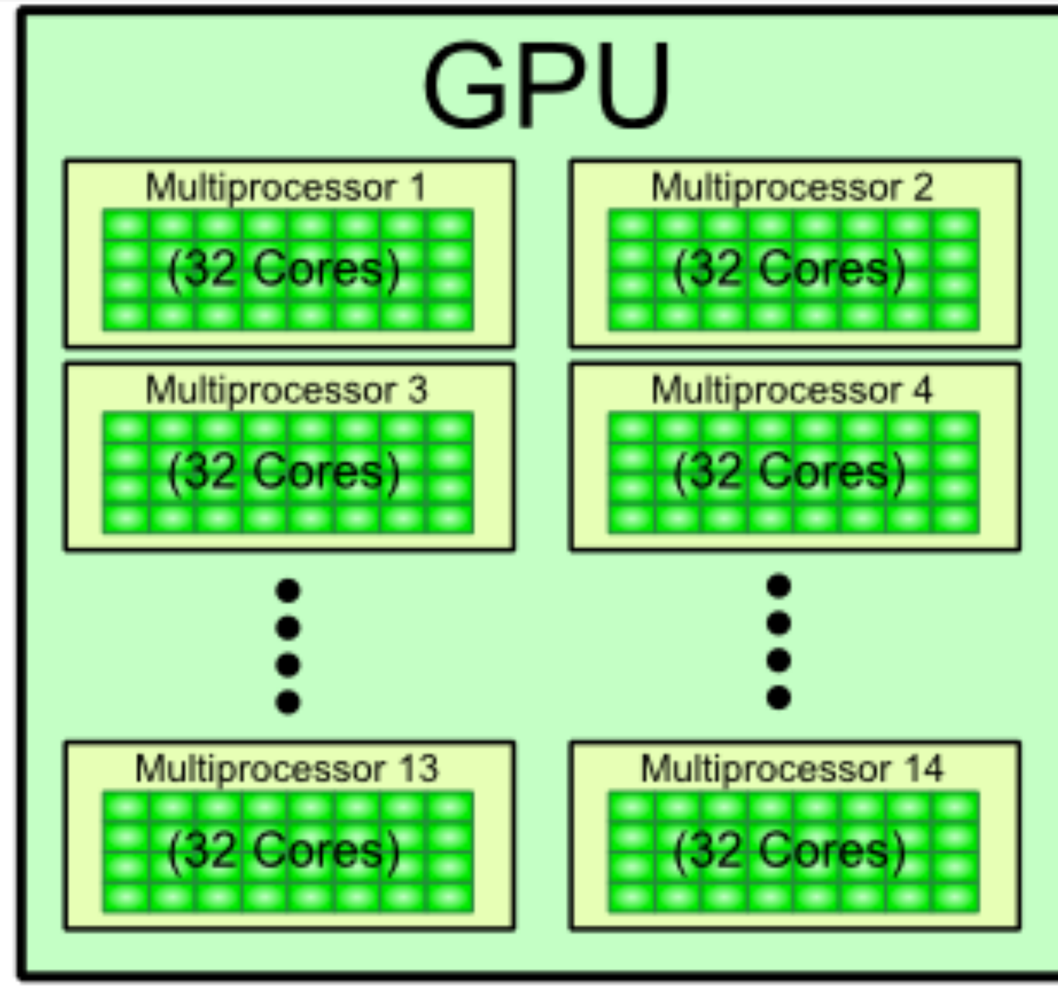
2006 Breakthrough

- More data
- **Faster hardware: GPU's, multi-core CPU's**
- Working ideas on how to train deep architectures

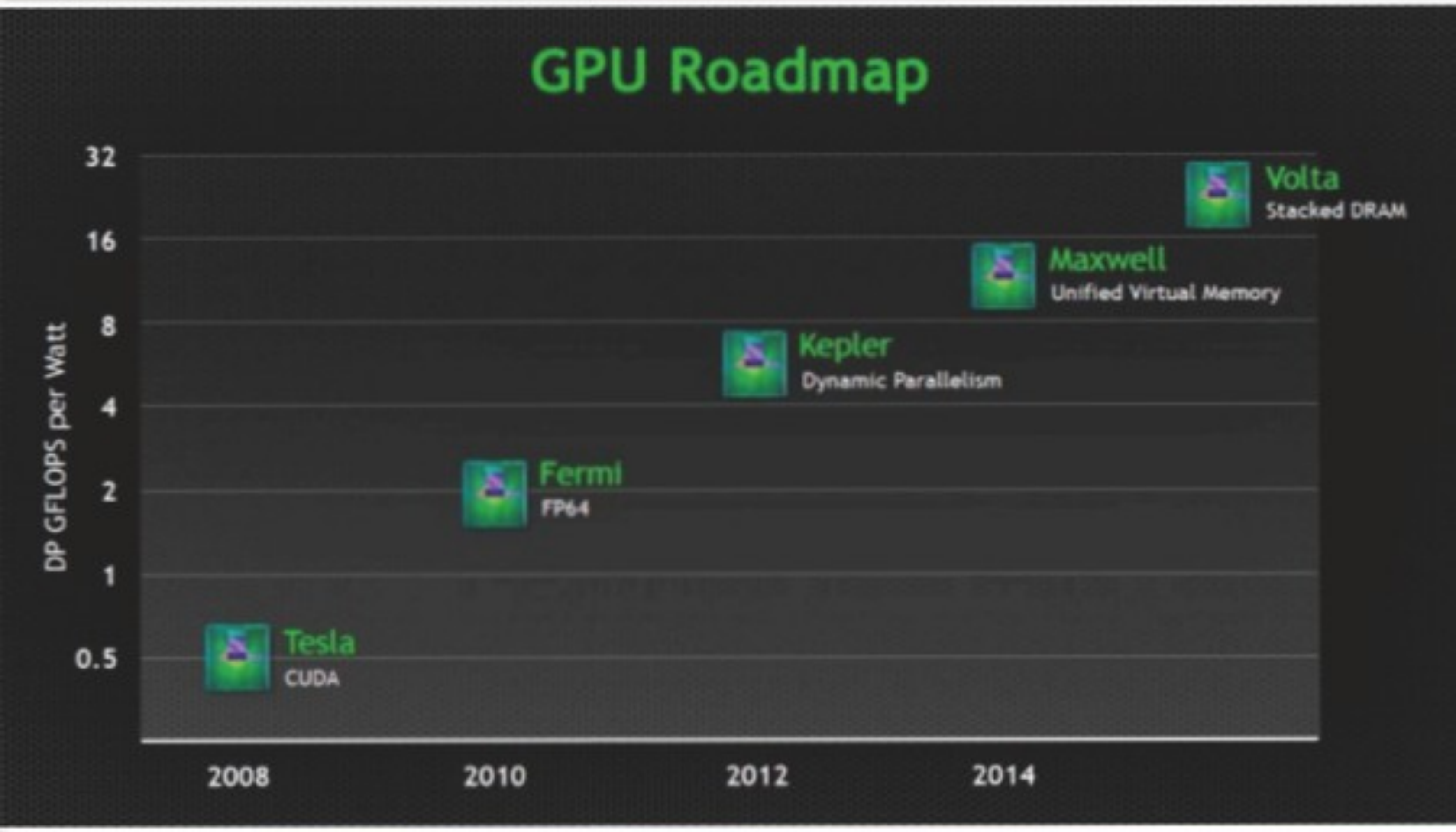
2006 Breakthrough

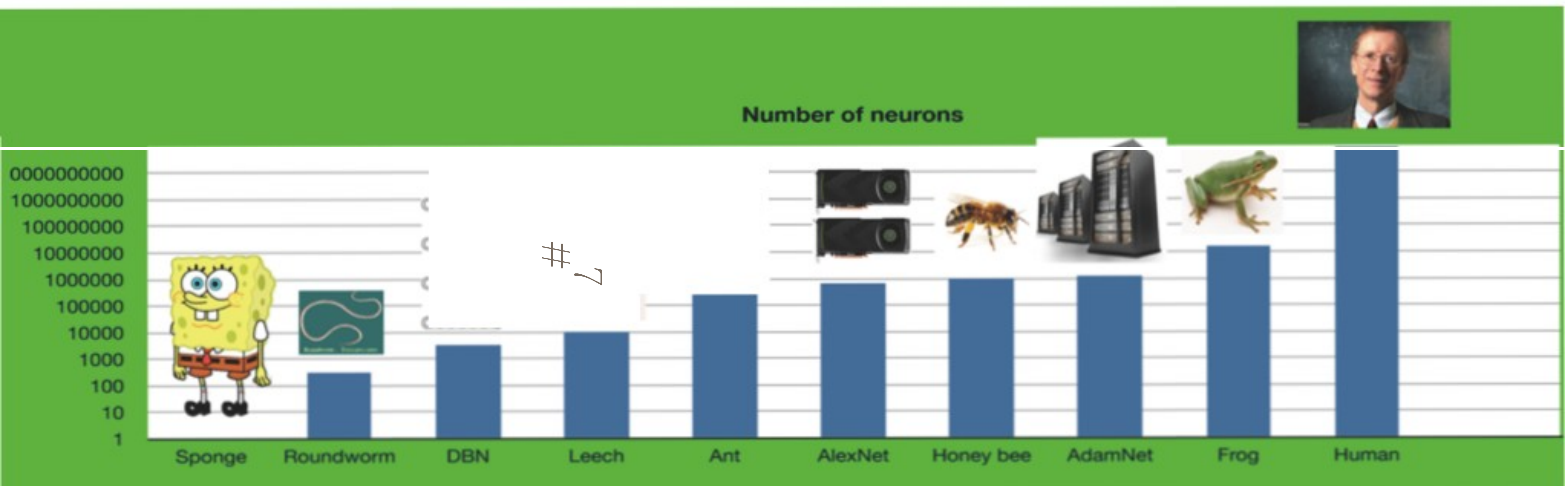


vs



Rise of Raw Computation Power





2006 Breakthrough

- More data
- Faster hardware: GPU's, multi-core CPU's
- **Working ideas on how to train deep architectures**

2006 Breakthrough

Stacked Restricted Boltzman Machines* (RBM)

Hinton, G. E, Osindero, S., and Teh, Y. W. (2006).
A fast learning algorithm for deep belief nets.
Neural Computation, 18:1527-1554.

Stacked Autoencoders (AE)

Bengio, Y., Lamblin, P., Popovici, P., Larochelle, H. (2007).
Greedy Layer-Wise Training of Deep Networks,
Advances in Neural Information Processing Systems 19

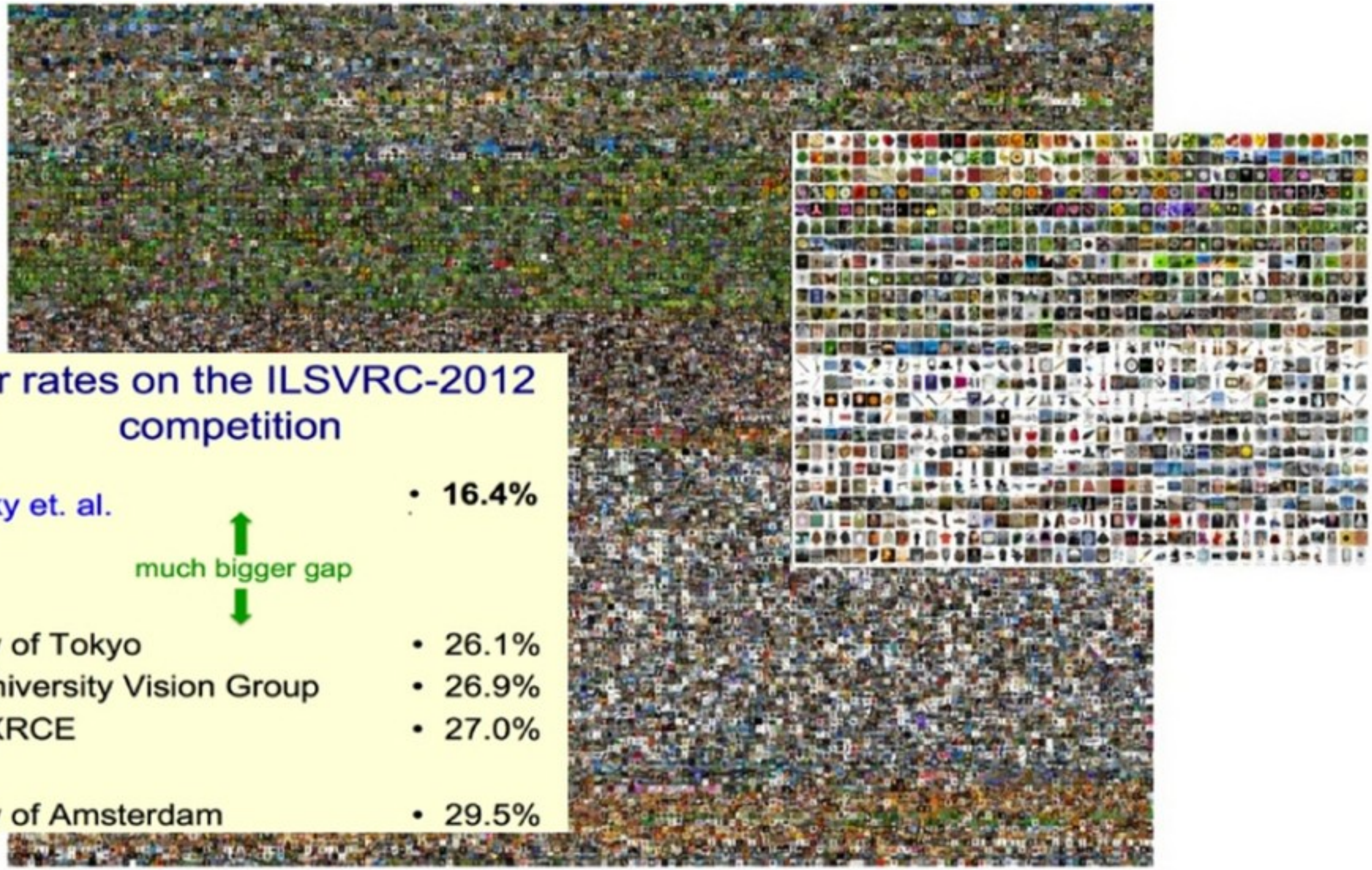
* *Called Deep Belief Networks (DBN)*

Deep Learning for the Win!

Impact on Computer Vision

ImageNet Challenge 2012

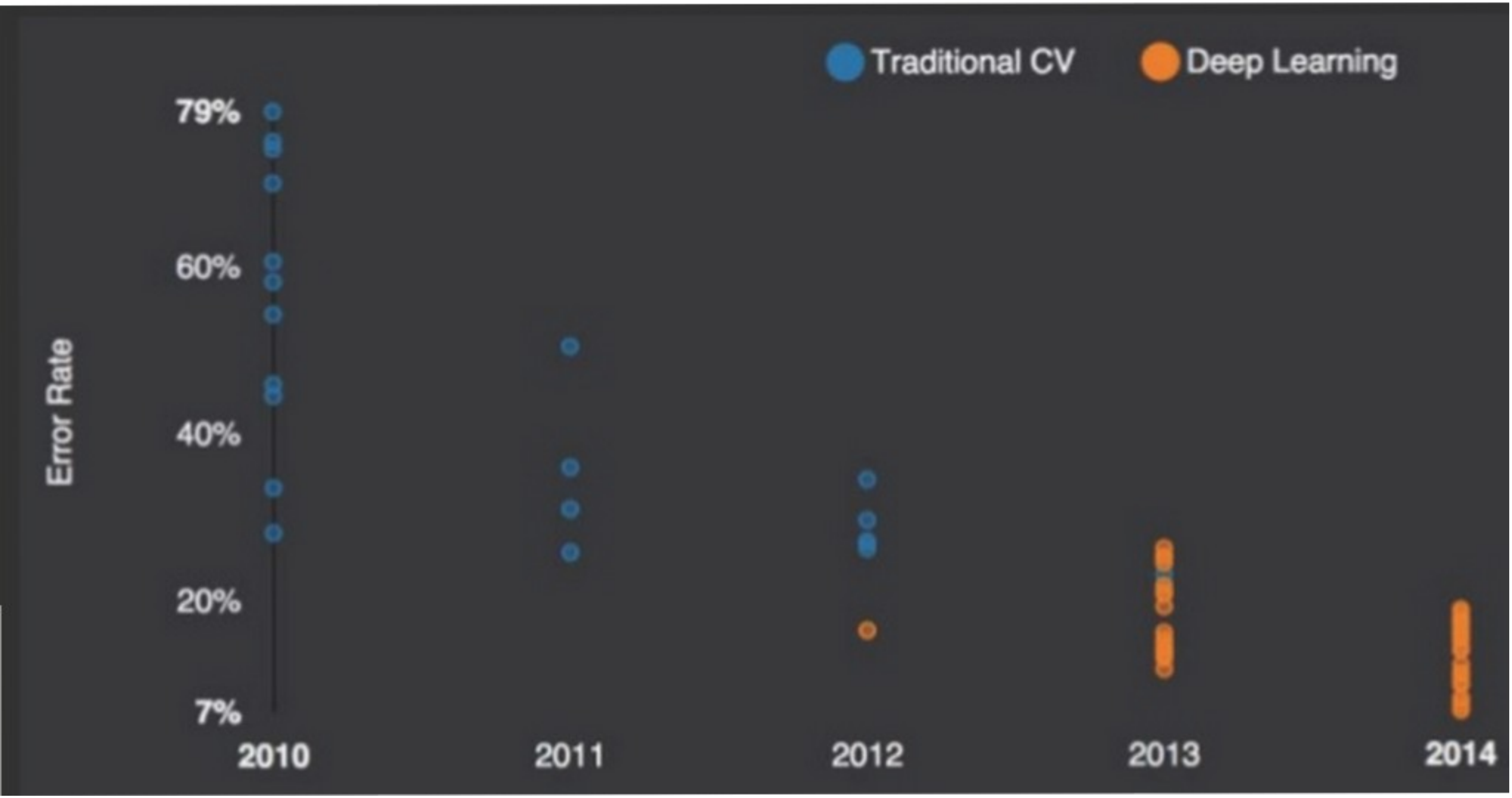
- 1.2M images with 1000 object categories



Error rates on the ILSVRC-2012 competition

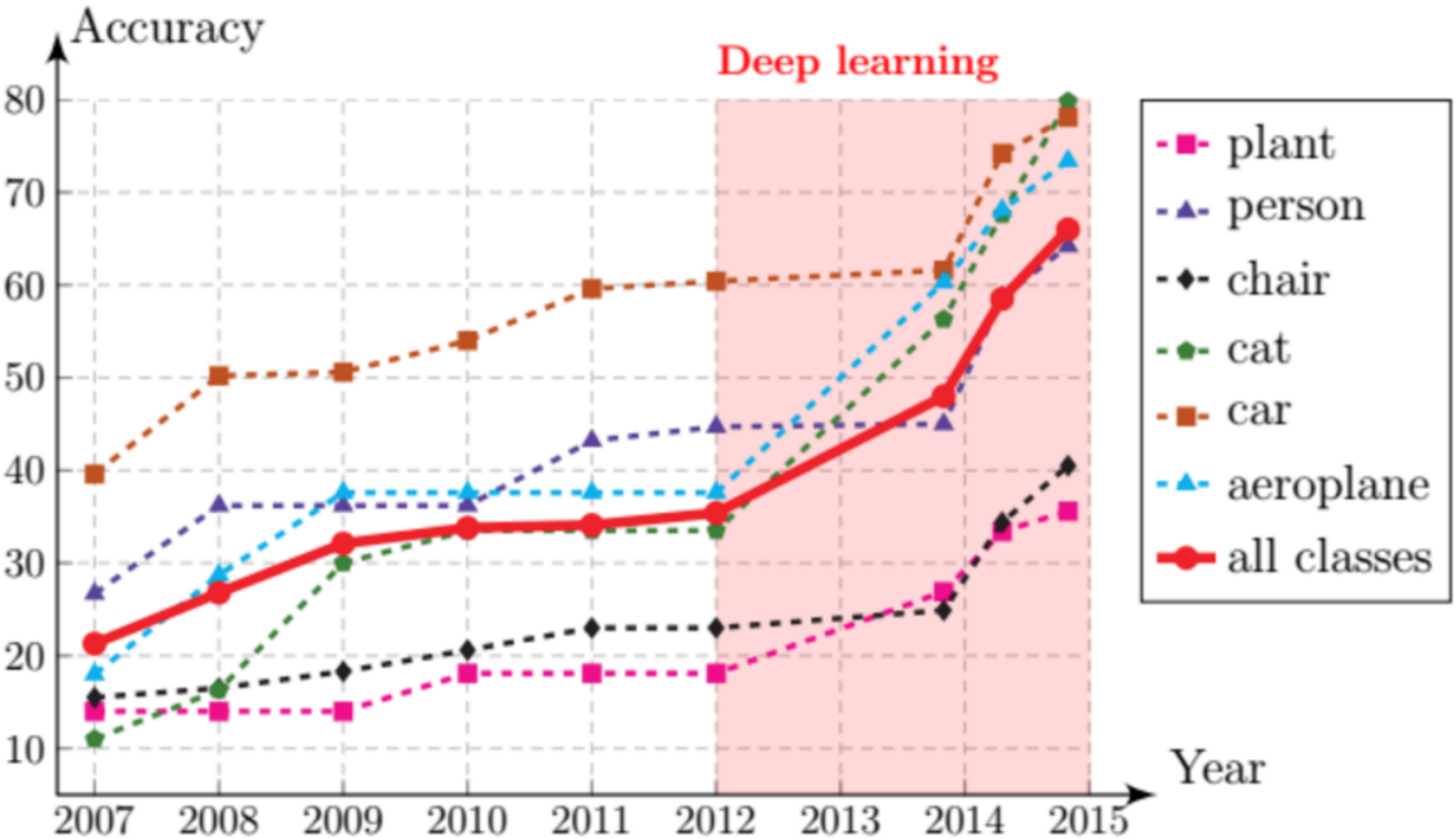
• Krizhevsky et. al.	• 16.4%
much bigger gap	
• University of Tokyo	• 26.1%
• Oxford University Vision Group	• 26.9%
• INRIA + XRCE	• 27.0%
• University of Amsterdam	• 29.5%

Impact on Computer Vision



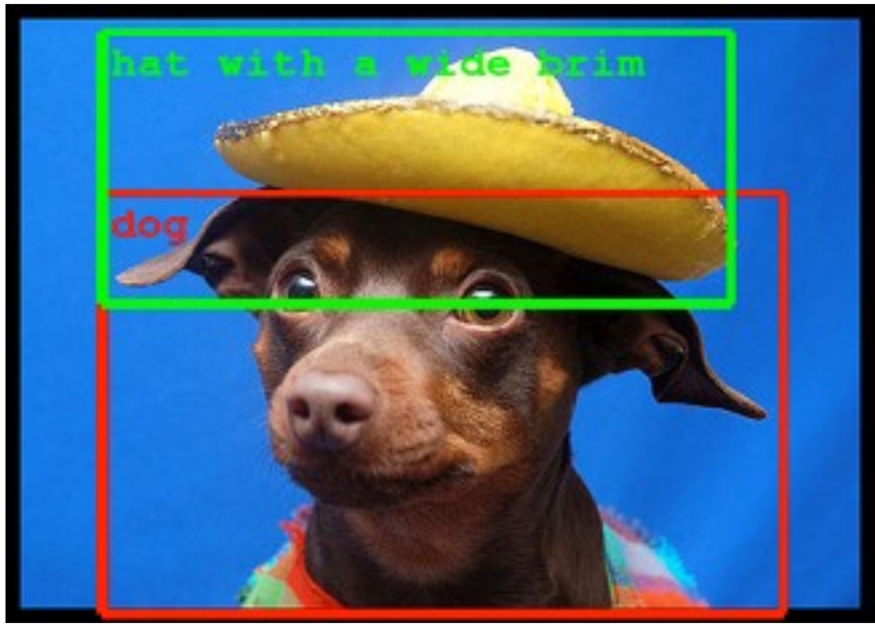
(from Clarifai)

Impact on Computer Vision



Classification results on ImageNet 2012

Team	Year	Place	Error (top-5)	Uses external data
SuperVision	2012	-	16.4%	no
SuperVision	2012	1st	15.3%	ImageNet 22k
Clarifai	2013	-	11.7%	no
Clarifai	2013	1st	11.2%	ImageNet 22k
MSRA	2014	3rd	7.35%	no
VGG	2014	2nd	7.32%	no
GoogLeNet	2014	1st	6.67%	no



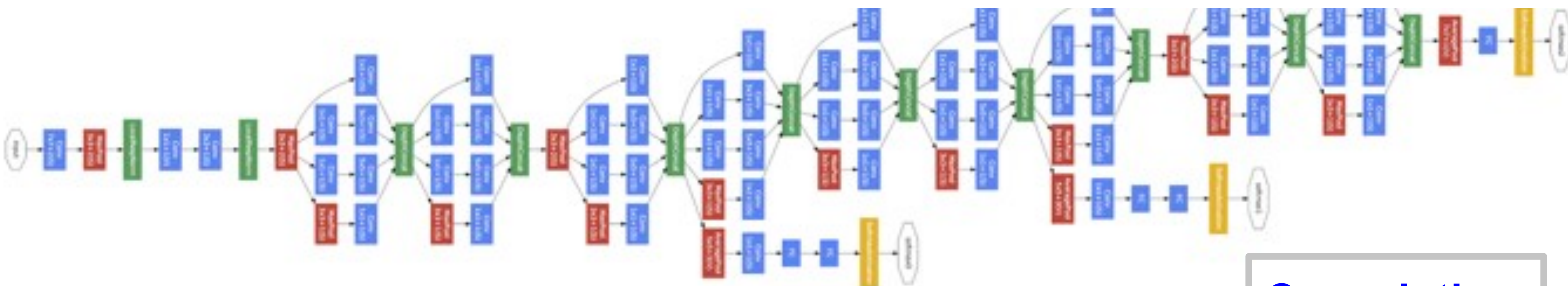
Detection results

Team	Year	Place	mAP	external data	ensemble	contextual model	approach
UvA-Euvision	2013	1st	22.6%	none	?	yes	Fisher vectors
Deep Insight	2014	3rd	40.5%	ILSVRC12 Classification + Localization	3 models	yes	ConvNet
CUHK DeepID-Net	2014	2nd	40.7%	ILSVRC12 Classification + Localization	?	no	ConvNet
GoogLeNet	2014	1st	43.9%	ILSVRC12 Classification	6 models	no	ConvNet

source: Szegedy et al. Going deeper with convolutions (GoogLeNet), ILSVRC2014, 19 Sep 2014

Winners of:
Large Scale Visual Recognition Challenge 2014 (ILSVRC2014)
19 September 2014

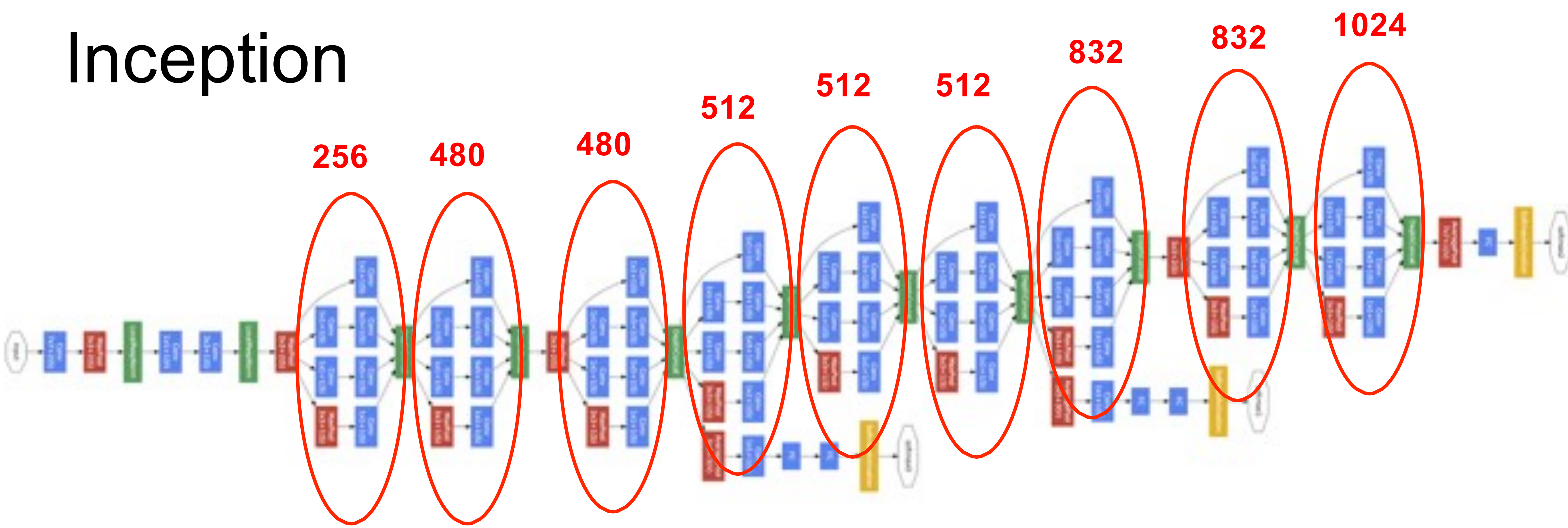
GoogLeNet



Convolution
Pooling
Softmax
Other

source: Szegedy et al. Going deeper with convolutions (GoogLeNet), ILSVRC2014, 19 Sep 2014

Inception



Width of inception modules ranges from 256 filters (in early modules) to 1024 in top inception modules.

Can remove fully connected layers on top completely

Number of parameters is reduced to 5 million

Computional cost is increased by less than 2X compared to Krizhevsky's network. (<1.5Bn operations/evaluation)

source: Szegedy et al. Going deeper with convolutions (GoogLeNet), ILSVRC2014, 19 Sep 2014

Impact on Computer Vision

Latest State of the Art:

Team	Date	Top-5 test error
GoogLeNet	2014	6.66%
Deep Image	01/12/2015	5.98%
Deep Image	02/05/2015	5.33%
Microsoft	02/05/2015	4.94%
Google	03/02/2015	**4.82%**
Deep Image	03/17/2015	4.83%

Computer Vision: Current State of the Art

Impact on Audio Processing

First public Breakthrough with Deep Learning in 2010

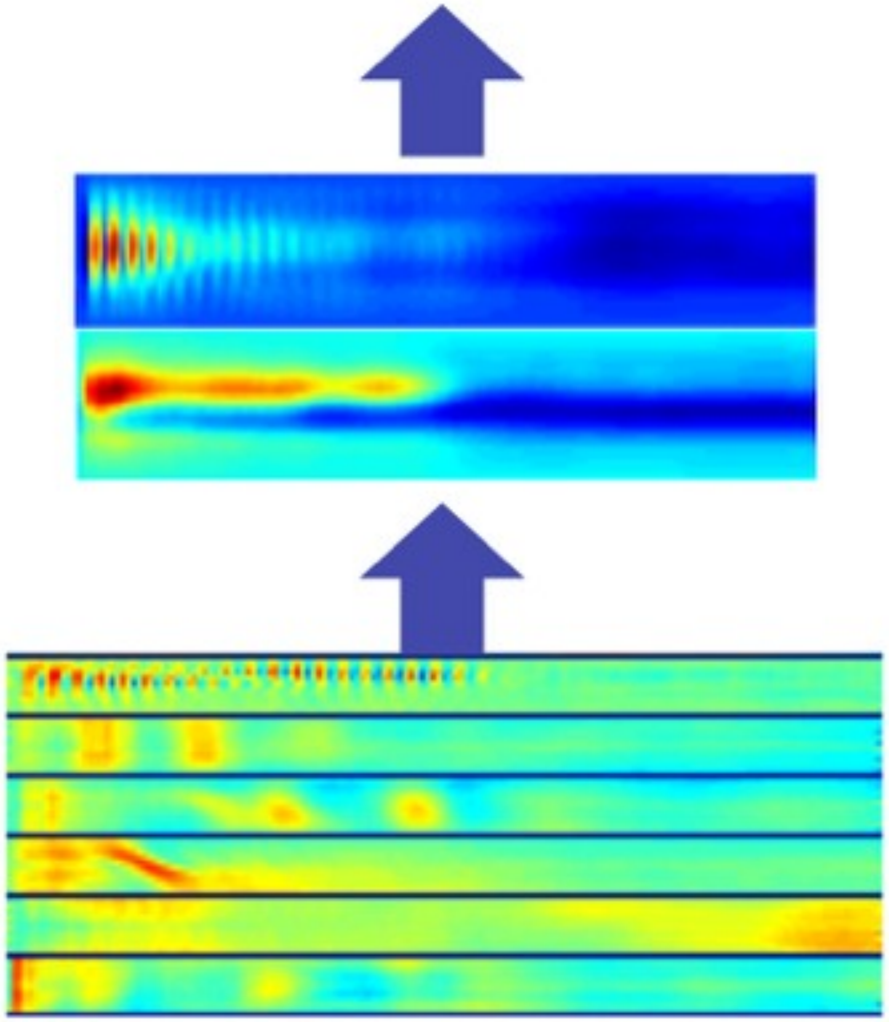
Dahl et al. (2010)

IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 1, JANUARY 2012

Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition

George E. Dahl, Dong Yu, *Senior Member, IEEE*, Li Deng, *Fellow, IEEE*, and Alex Acero, *Fellow, IEEE*

Phonemes/Words



Acoustic model	Recog \ WER	RT03S FSH	Hub5 SWB
Traditional features	1-pass −adapt	**27.4**	**23.6**
Deep Learning	1-pass −adapt	**18.5**	**16.1**

Impact on Audio Processing

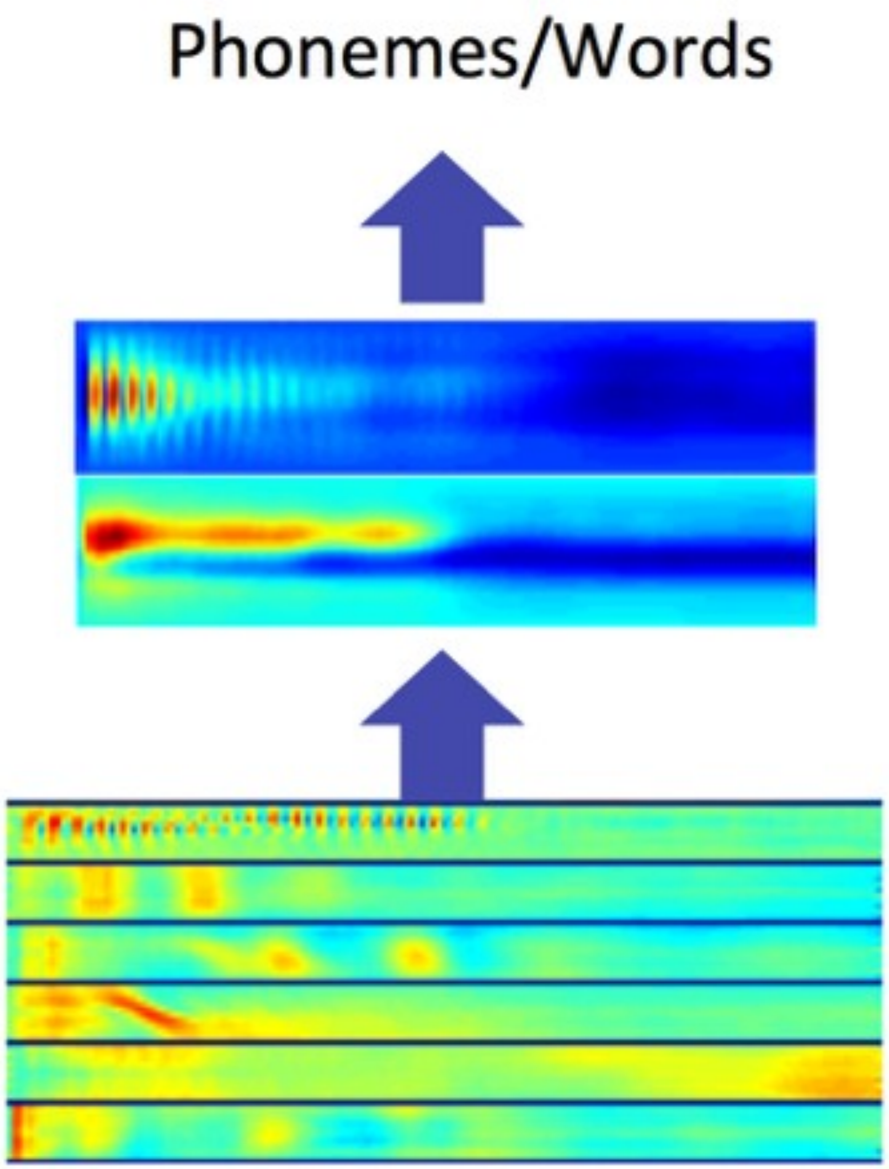
First public Breakthrough with Deep Learning in 2010

Dahl et al. (2010)

IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 20, NO. 1, JANUARY 2012

Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition

George E. Dahl, Dong Yu, *Senior Member, IEEE*, Li Deng, *Fellow, IEEE*, and Alex Acero, *Fellow, IEEE*

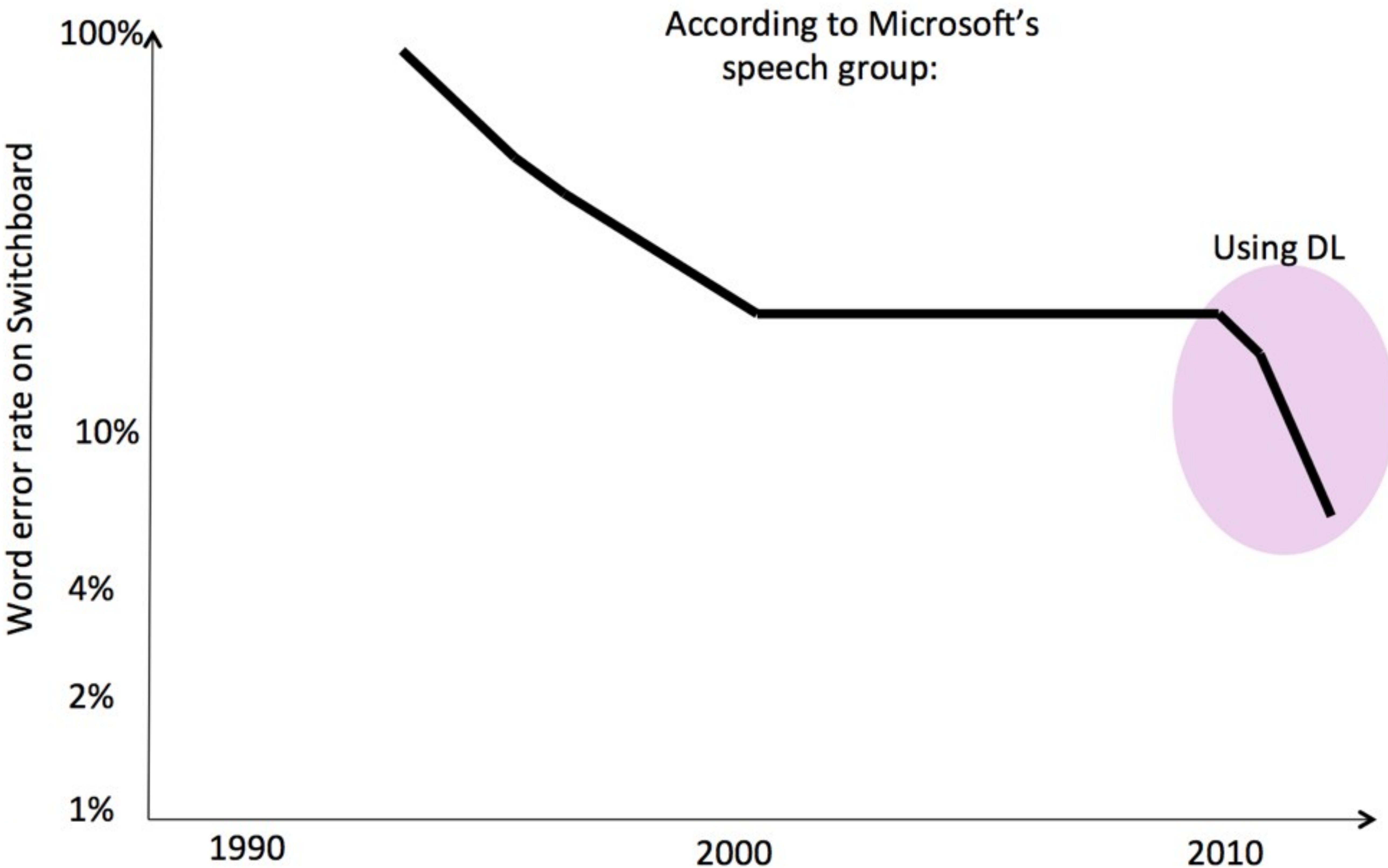


Acoustic model	Recog \ WER	RT03S FSH	Hub5 SWB
Traditional features	1-pass −adapt	27.4	23.6
Deep Learning	1-pass −adapt	18.5	16.1

-33%! -32%!

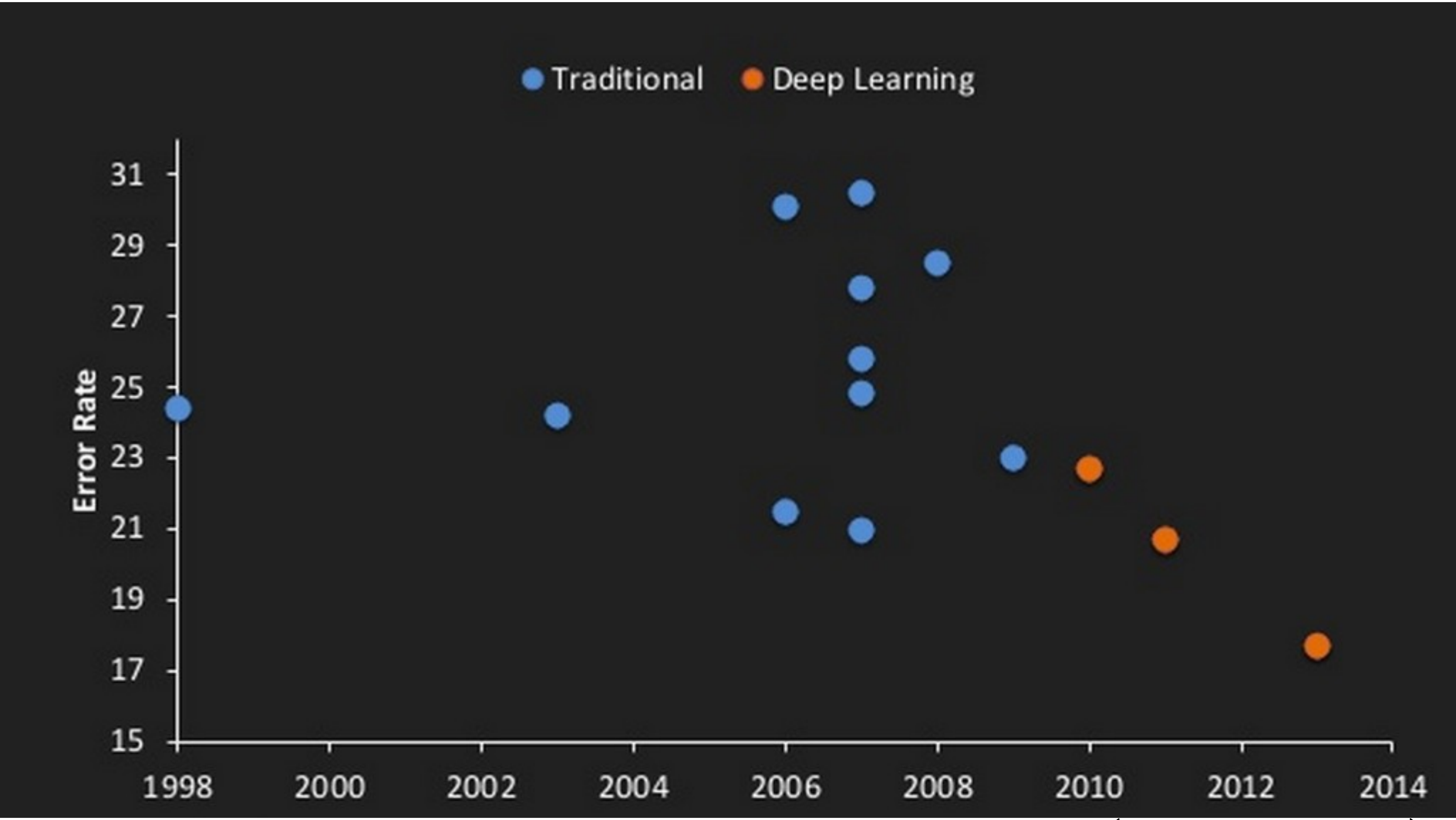
Impact on Audio Processing

Speech Recognition



Impact on Audio Processing

TIMIT Speech Recognition



(from: Clarifai)

Impact on Audio Processing

Deep Speech: Scaling up end-to-end speech recognition

Awni Hannun,* Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, Andrew Y. Ng

Baidu Research – Silicon Valley AI Lab

Model	SWB	CH	Full
Vesely et al. (GMM-HMM BMMI) [44]	18.6	33.0	25.8
Vesely et al. (DNN-HMM sMBR) [44]	12.6	24.1	18.4
Maas et al. (DNN-HMM SWB) [28]	14.6	26.3	20.5
Maas et al. (DNN-HMM FSH) [28]	16.0	23.7	19.9
Seide et al. (CD-DNN) [39]	16.1	n/a	n/a
Kingsbury et al. (DNN-HMM sMBR HF) [22]	13.3	n/a	n/a
Sainath et al. (CNN-HMM) [36]	11.5	n/a	n/a
Soltau et al. (MLP/CNN+I-Vector) [40]	**10.4**	n/a	n/a
Deep Speech SWB	20.0	31.8	25.9
Deep Speech SWB + FSH	12.6	**19.3**	**16.0**

Impact on Natural Language Processing

	POS WSJ (acc.)	NER CoNLL (F1)
State-of-the-art*	97.24	89.31
Supervised NN	96.37	81.47
NN with Brown clusters	96.92	87.15
Word vector pre-training followed by supervised NN**	97.20	88.87
+ hand-crafted features***	97.29	89.59

Pos: Toutanova et al. 2003)

Ner: Ando & Zhang 2005

C&W 2011

C&W 2011

Impact on Natural Language Processing

Feature	NER
Current Word	✓
Previous Word	✓
Next Word	✓
Current Word Character n-gram	all
Current POS Tag	✓
Surrounding POS Tag Sequence	✓
Current Word Shape	✓
Surrounding Word Shape Sequence	✓
Presence of Word in Left Window	size 4
Presence of Word in Right Window	size 4

Named Entity Recognition:

Model	Dev	Test	ACE	MUC7
Discrete	91.0	85.4	77.4	73.4
SVD	90.8	85.7	77.3	73.7
SVD-S	91.0	85.5	77.6	74.3
SVD-L	90.5	84.8	73.6	71.5
HPCA	92.6	**88.7**	81.7	80.7
HSMN	90.5	85.7	78.7	74.7
CW	92.2	87.4	81.7	80.2
CBOW	93.1	88.2	82.2	81.1
GloVe	**93.2**	88.3	**82.9**	**82.2**

Deep Learning: Who's to blame?

Deep Learning: Who’s to blame?

Deep Learning: Why?

Deep Architectures can be representationally efficient

- Fewer computational units for same function

Deep Representations might allow for a hierarchy or representation

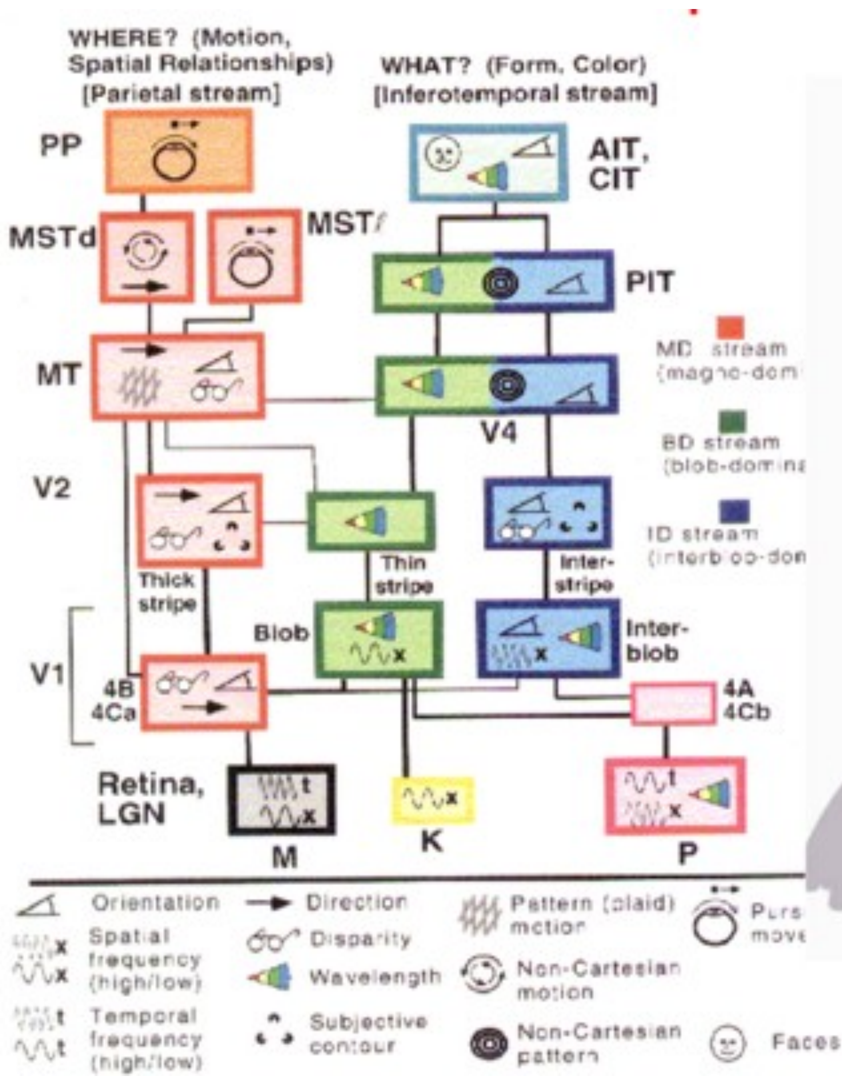
- Allows non-local
- generalisation

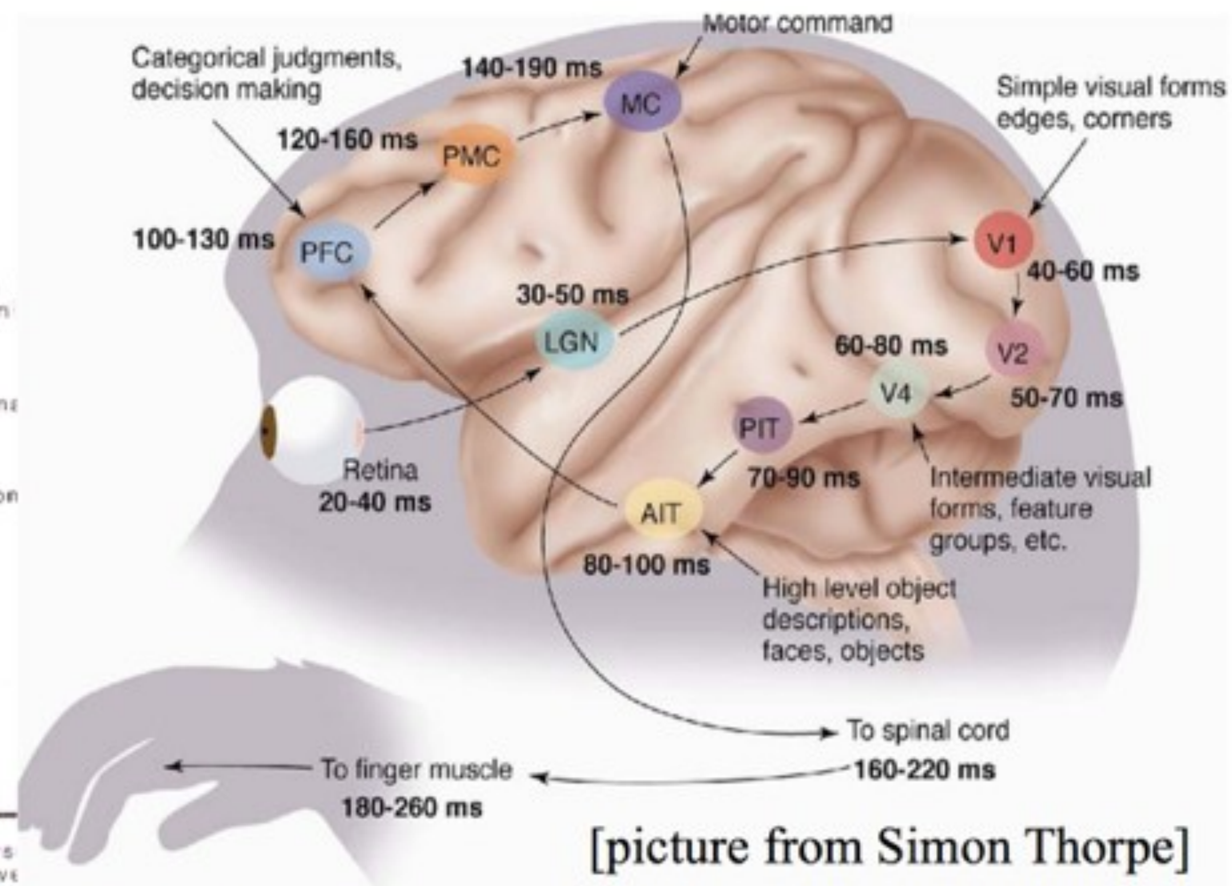
  Comprehensibility

Multiple levels of latent variables allow combinatorial sharing of statistical strength

Deep Learning = Brain “inspired”

Audio/Visual Cortex has multiple stages == Hierarchical

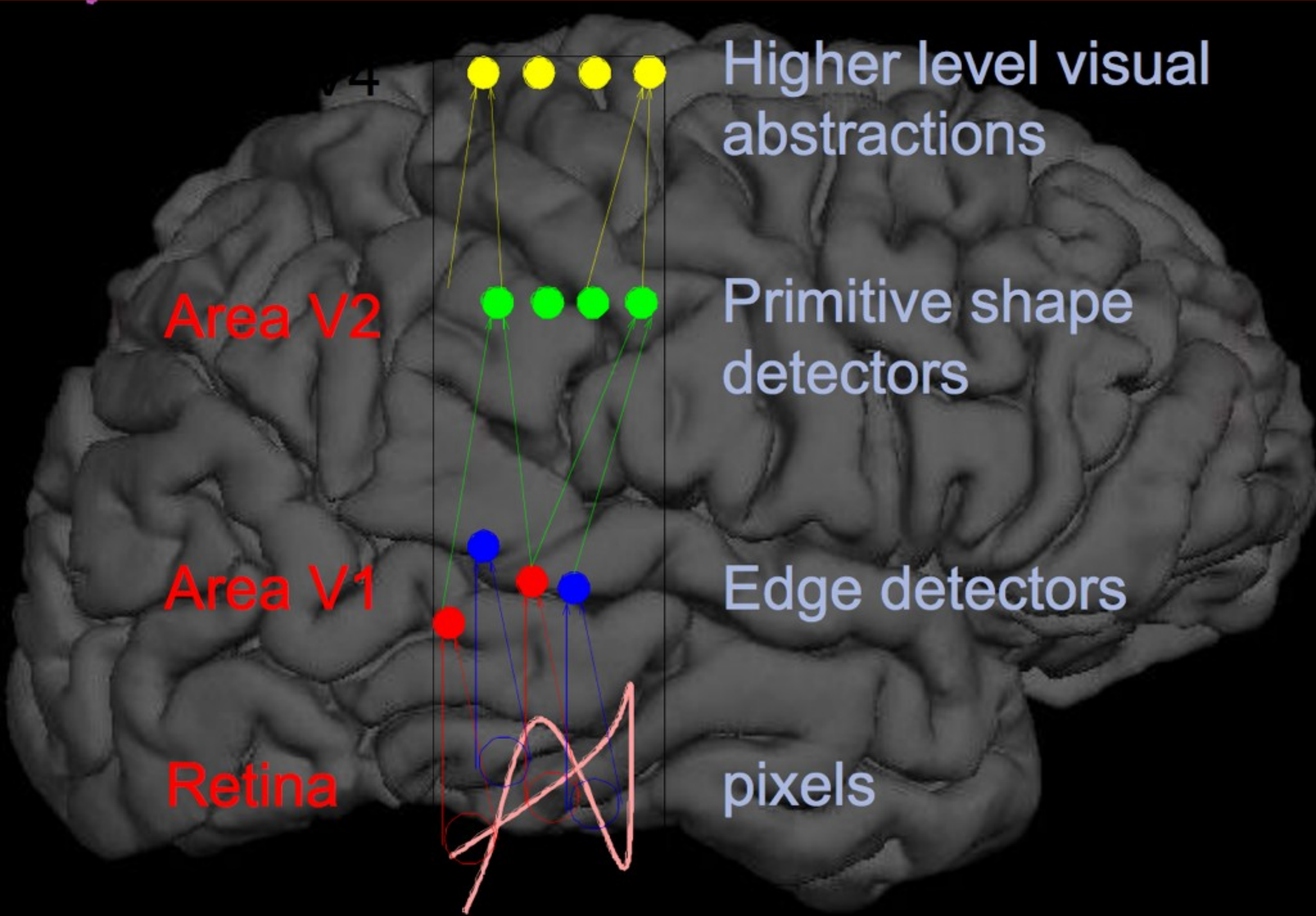




[picture from Simon Thorpe]

[Gallant & Van Essen]

Different Levels of Abstraction

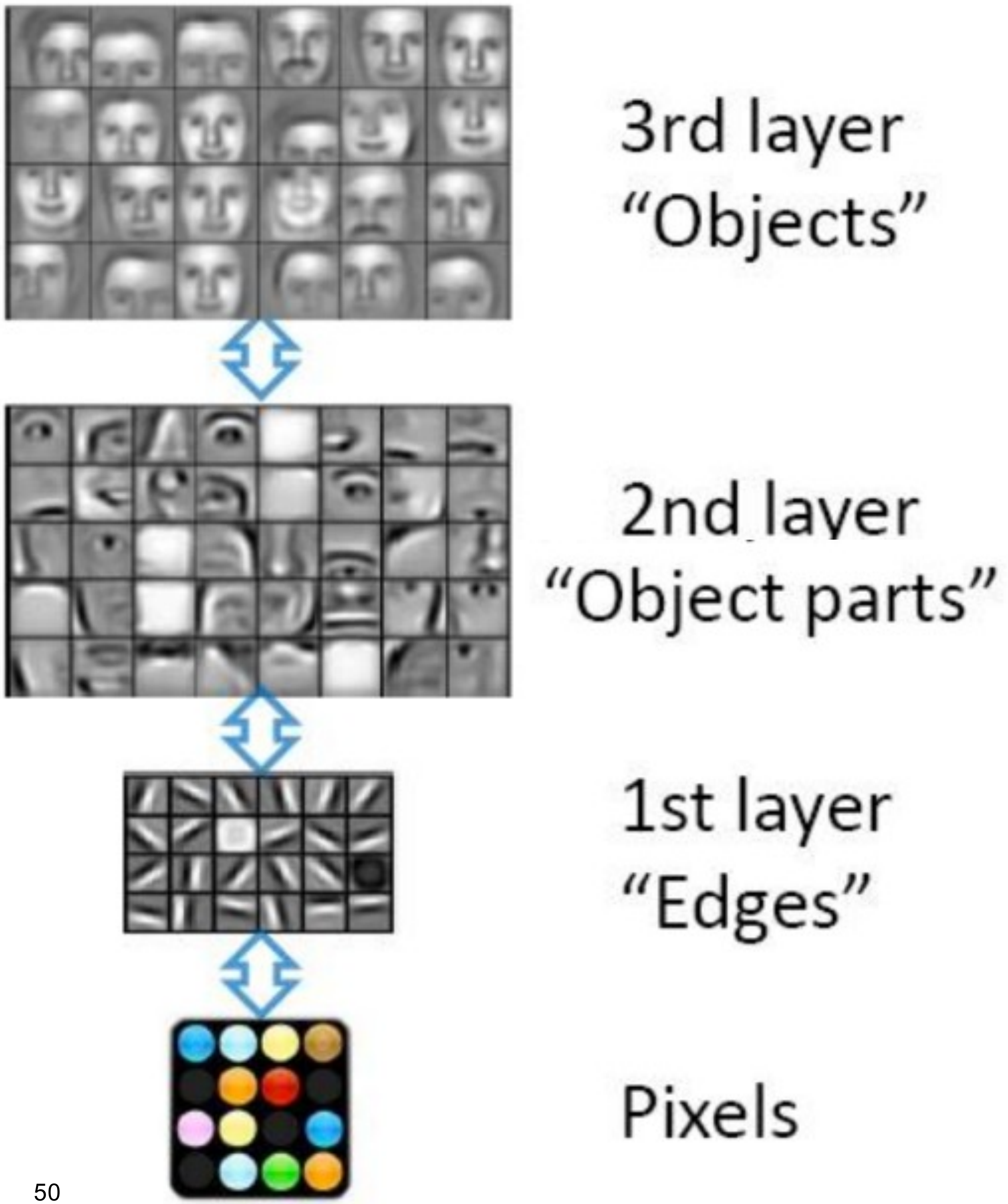


Different Levels of Abstraction

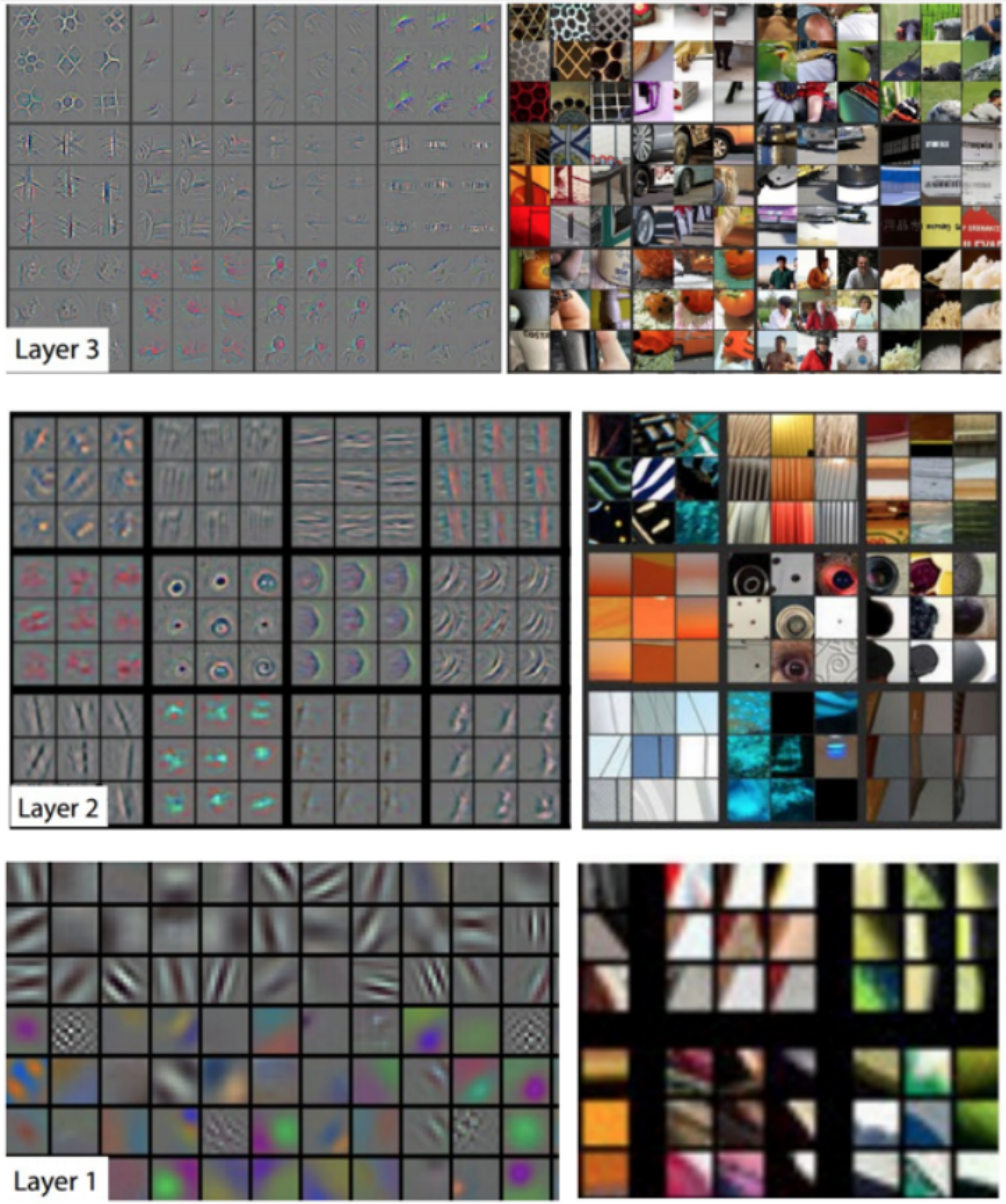
Hierarchical Learning

- Natural progression from low level to high level structure as seen in natural complexity
- A good lower level representation can be used for many distinct tasks

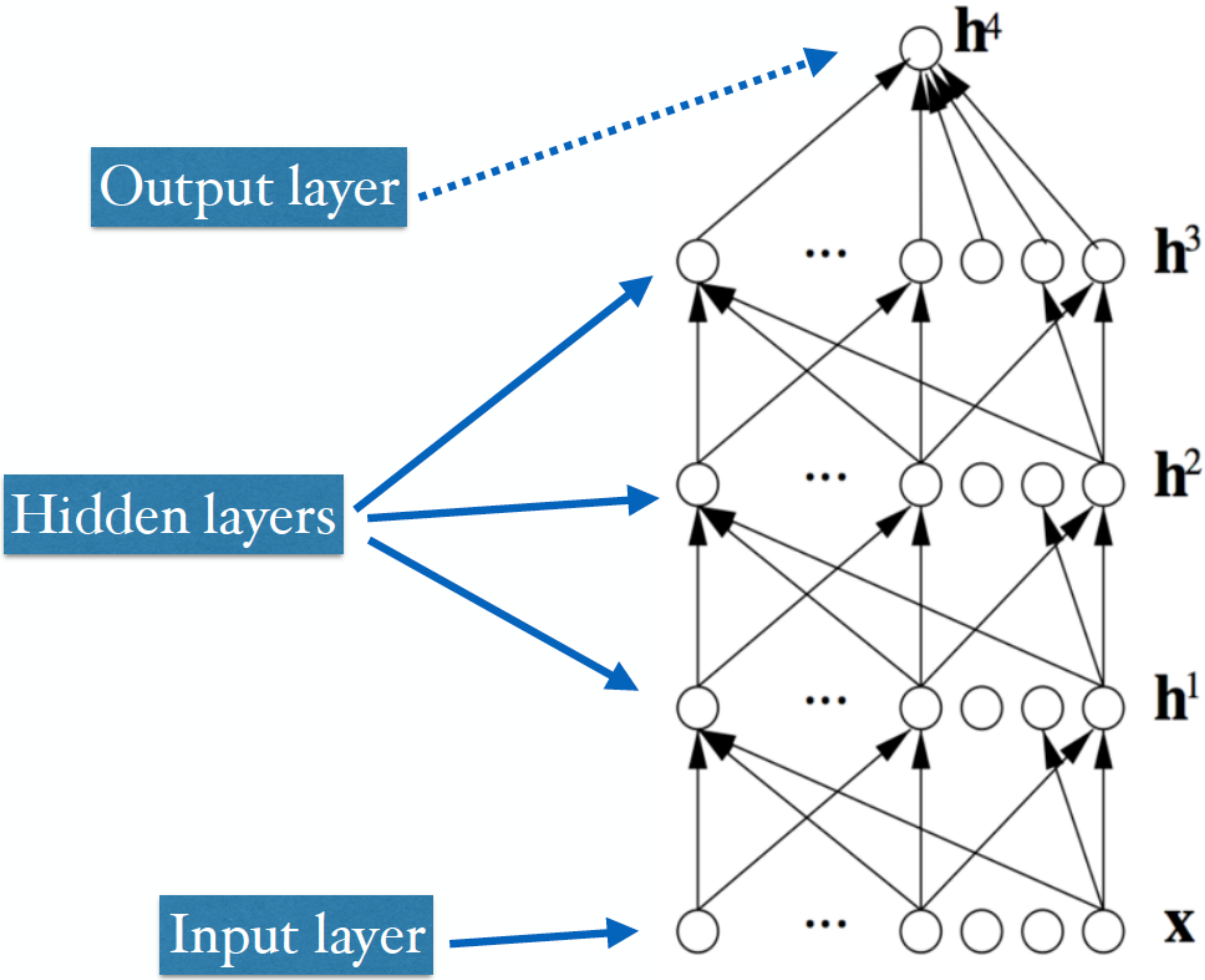
Feature Representation



Different Levels of Abstraction

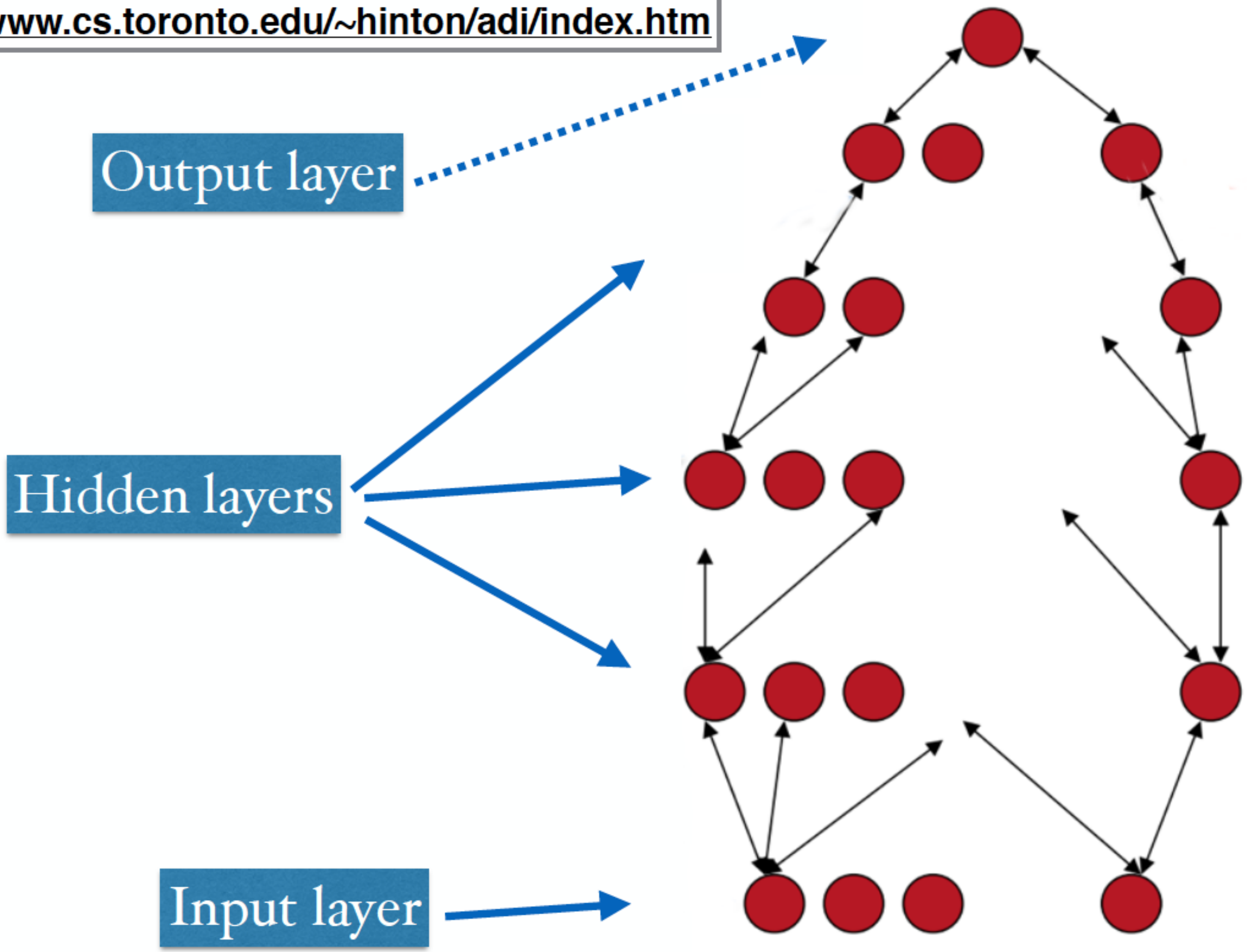


Classic Deep Architecture



Modern Deep Architecture

movie time:
http://www.cs.toronto.edu/~hinton/adi/index.htm



Why go Deep ?

Hierarchies

Black Box

Distributed

Training Time

Efficient

Sharin

Much Data

g Generalizatio

Unsupervised*

Major PWNAGE!


No More Handcrafted Features !

Why Deep Learning ?

- Manually designed features are often over-specified, incomplete and take a long time to design and validate
- **Learned Features** are easy to adapt, fast to learn

- Deep learning provides a very flexible, (almost?) universal, learnable framework for **representing** world, visual and linguistic information.

- Deep learning can learn **unsupervised** (from raw text/audio/images/whatever content) and **supervised** (with specific labels like positive/negative)

[Kudos to Richard Socher, for this eloquent summary :)]

Deep Learning: Future Developments

Currently an explosion of developments

- Hessian-Free networks (2010)
- Long Short Term Memory (2011)
- Large Convolutional nets, max-pooling (2011)
- Nesterov's Gradient Descent (2013)

Currently state of the art but...

- No way of doing logical inference (extrapolation)
- No easy integration of abstract knowledge
- Hypothetic space bias might not conform with reality

Wanna Play ? General Deep Learning

- TensorFlow – Google Open Source. https://www.tensorflow.org/
  - A playground for neural network
- Theano - CPU/GPU symbolic expression compiler in python (from LISA lab at University of Montreal). http://deeplearning.net/software/theano/
- Torch - Matlab-like environment for state-of-the-art machine learning algorithms in lua (from Ronan Collobert, Clement Farabet and Koray Kavukcuoglu) http://torch.ch/
- More info: http://deeplearning.net/software links/

Wanna Play ? Computer Vision

- cuda-convnet2 (Alex Krizhevsky, Toronto) (c++/ CUDA, optimized for GTX 580) https://code.google.com/p/cuda-convnet2/
- Caffe (Berkeley) (Cuda/OpenCL, Theano, Python) http://caffe.berkeleyvision.org/
- OverFeat (NYU)
- http://cilvr.nyu.edu/doku.php?id=code:start