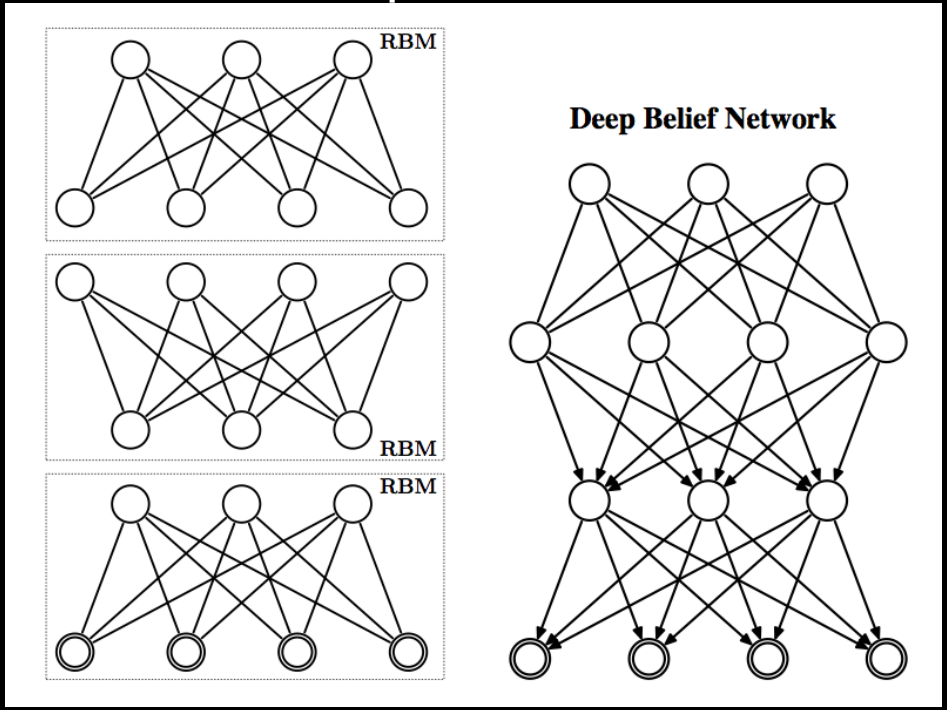
Introduction to Deep Learning

In Computer Vision



Dimitris Samaras
Stony Brook University

Slide credits & pointers

Iasonas Kokkinos
http://cvn.ecp.fr/personnel/iasonas/deeplearning.html
R. Fergus/K. Yu/M. A. Ranzatto (CVPR 12 Tutorial):
http://cs.nyu.edu/~fergus/tutorials/deep_learning_cvpr12/
S. Liu/S. Roth (ICCV 09 Tutorial): http://www.gris.informatik.tu-darmstadt.de/teaching/iccv2009/
E. Simoncelli (Class notes):
http://www.cns.nyu.edu/~eero/imrep-course/
A. Vedadi (BMVC14 Tutorial)
http://www.robots.ox.ac.uk/~vedaldi/assets/teach/vedaldi14bmvc-tutorial.pdf
Y. Bengio (Tutorial):
http://www.iro.umontreal.ca/~bengioy/talks/mlss-beijing.pdf
G. Papandreou, BASIS Tutorial
http://cvn.ecp.fr/personnel/iasonas/basis14/
Fei-Fei Li, Olga Russakovsky
http://ai.stanford.edu/~olga/slides/ImageNetAnalysis_bavm_10_5_13.pptx
R. Salakhudinov
http://www.cs.toronto.edu/~rsalakhu/kdd.html
IPAM summer school on deep learning:
http://www.ipam.ucla.edu/programs/summer-schools/graduate-summer-school-deep-learning-feature-learning/

Lecture outline

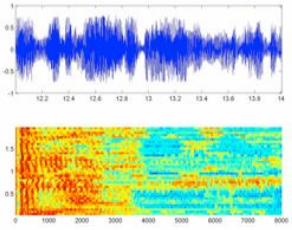
- Introduction to the class
- Flat models
  - Natural image statistics & probabilistic modelling
  - Sparse coding

Data: Big and Complicated

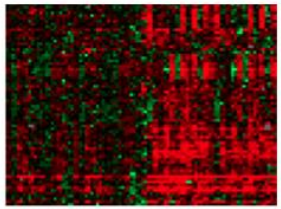
Images & Video

Text & Language

Speech & Audio

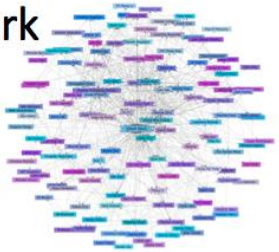


Gene Expression

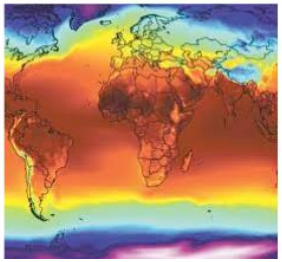


Product Recommendation

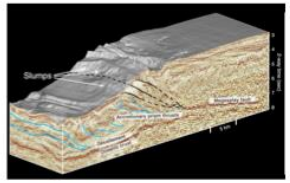
Relational Data/ Social Network



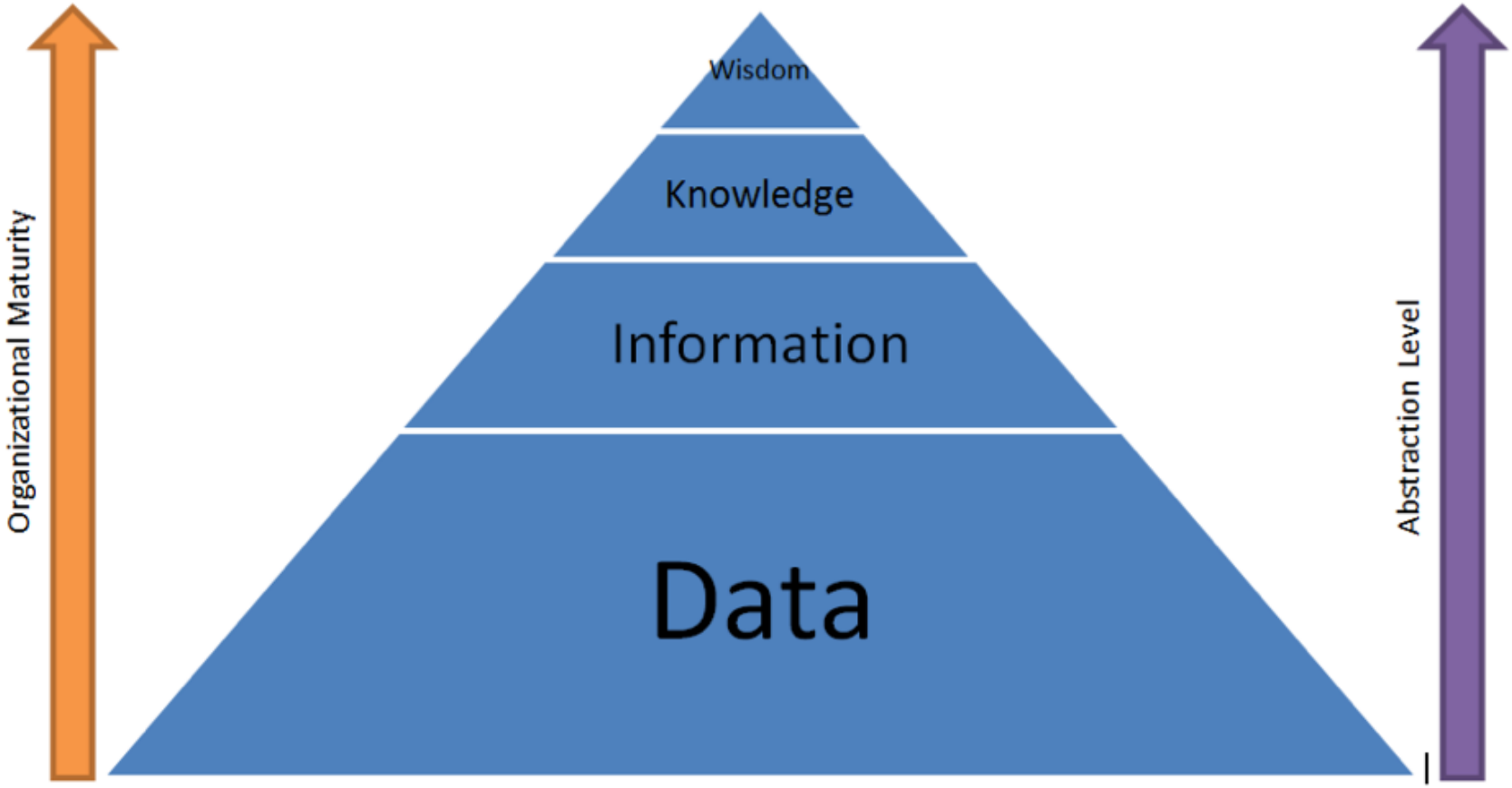
Climate Change



Geological Data



The data science quest



The computer vision quest

Backpack

Flute

Strawberry

Traffic light

Backpack

Matchstick

Sea lion

Bathing cap

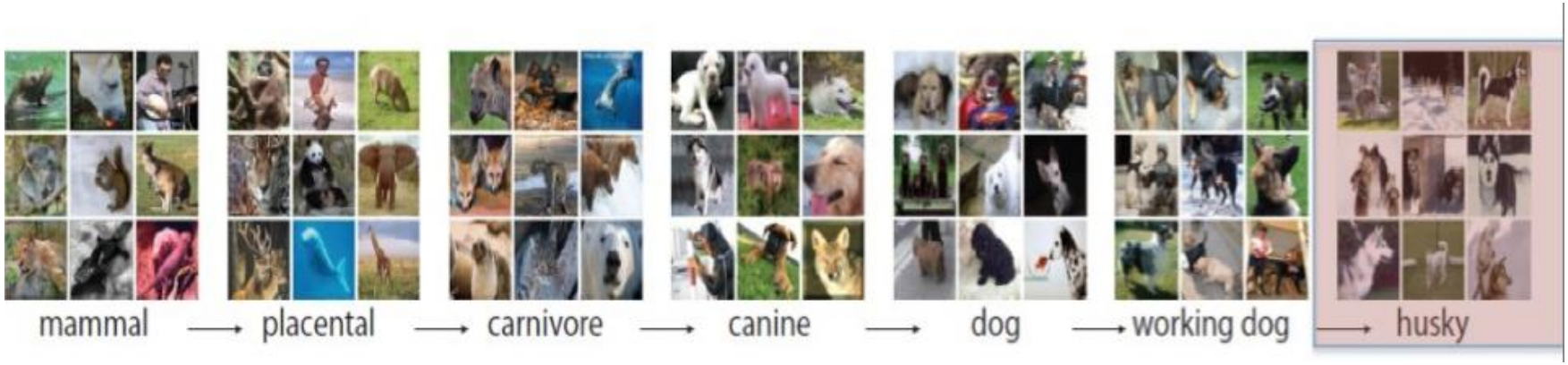
Racket


WAIT

Computer Vision Data: Big and Complicated

http://www.image-net.org/

IMAGENET



Deng et. al, ImageNet: A Large-Scale Hierarchical Image Database, CVPR'09

Computer Vision Data: Big and Complicated

http://www.image-net.org/

IMAGENET

Overall

- Total number of non-empty synsets: 21841
- Total number of images: 14,197,122
- Number of images with bounding box annotations: 1,034,908

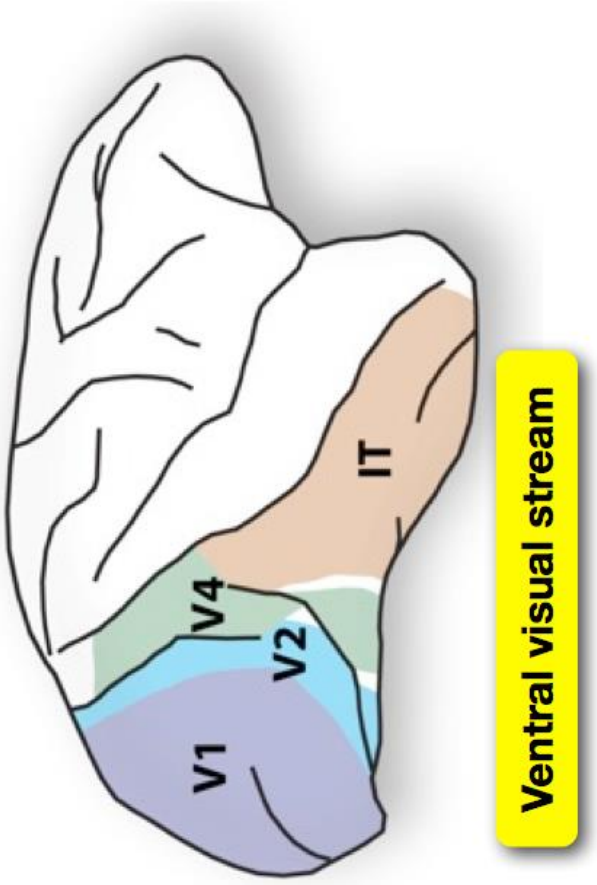
High level category	# synset (subcategories)	Avg # images per synset	Total # images
amphibian	94	591	56K
animal	3822	732	2799K
appliance	51	1164	59K
bird	856	949	812K
covering	946	819	774K
device	2385	675	1610K
fabric	262	690	181K
fish	566	494	280K
flower	462	735	339K
food	1495	670	1001K
fruit	309	607	188K
fungus	303	453	137K
furniture	187	1043	195K
geological formation	151	838	127K
invertebrate	728	573	417K
mammal	1138	821	934K
musical instrument	157	891	140K
plant	1666	600	999K
reptile	268	707	190K
sport	166	1207	200K
structure	1239	763	946K
tool	316	551	174K
tree	993	568	564K
utensil	86	912	78K
vegetable	176	764	135K
vehicle	481	778	374K
person	2035	468	952K

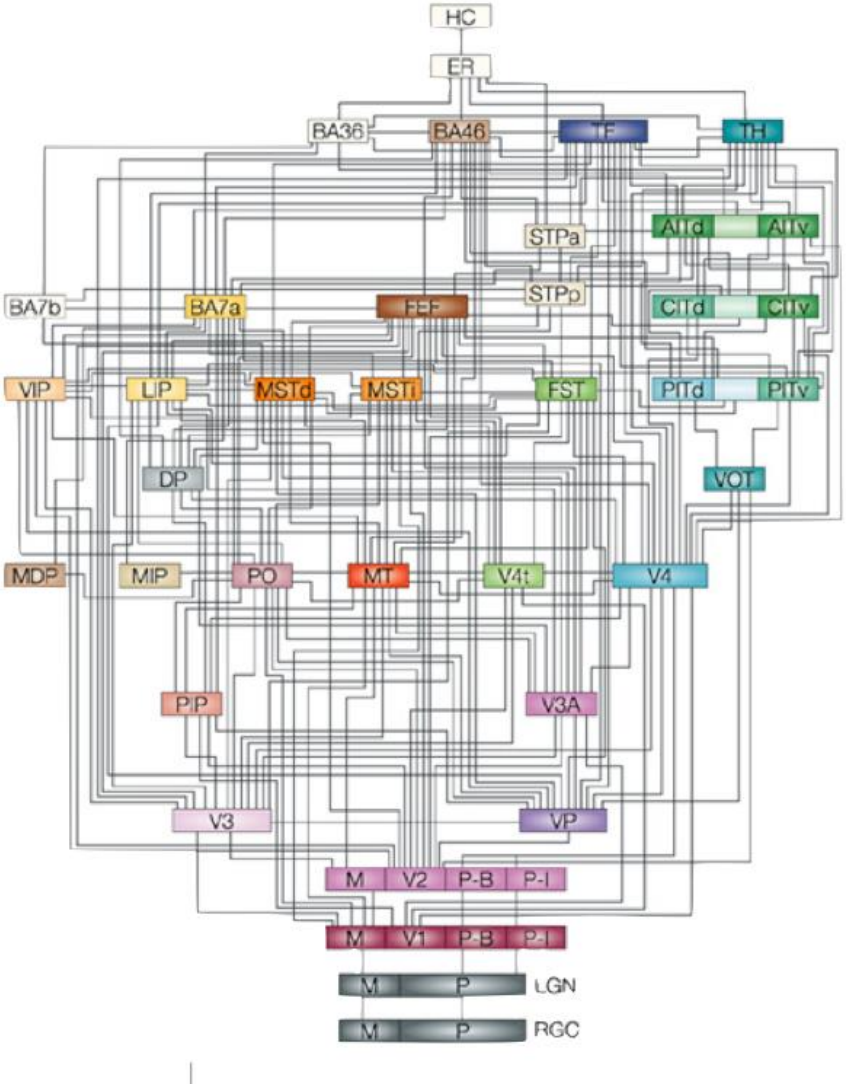
Deng et. al, ImageNet: A Large-Scale Hierarchical Image Database, CVPR'09

Computer Vision Data: Big and Complicated

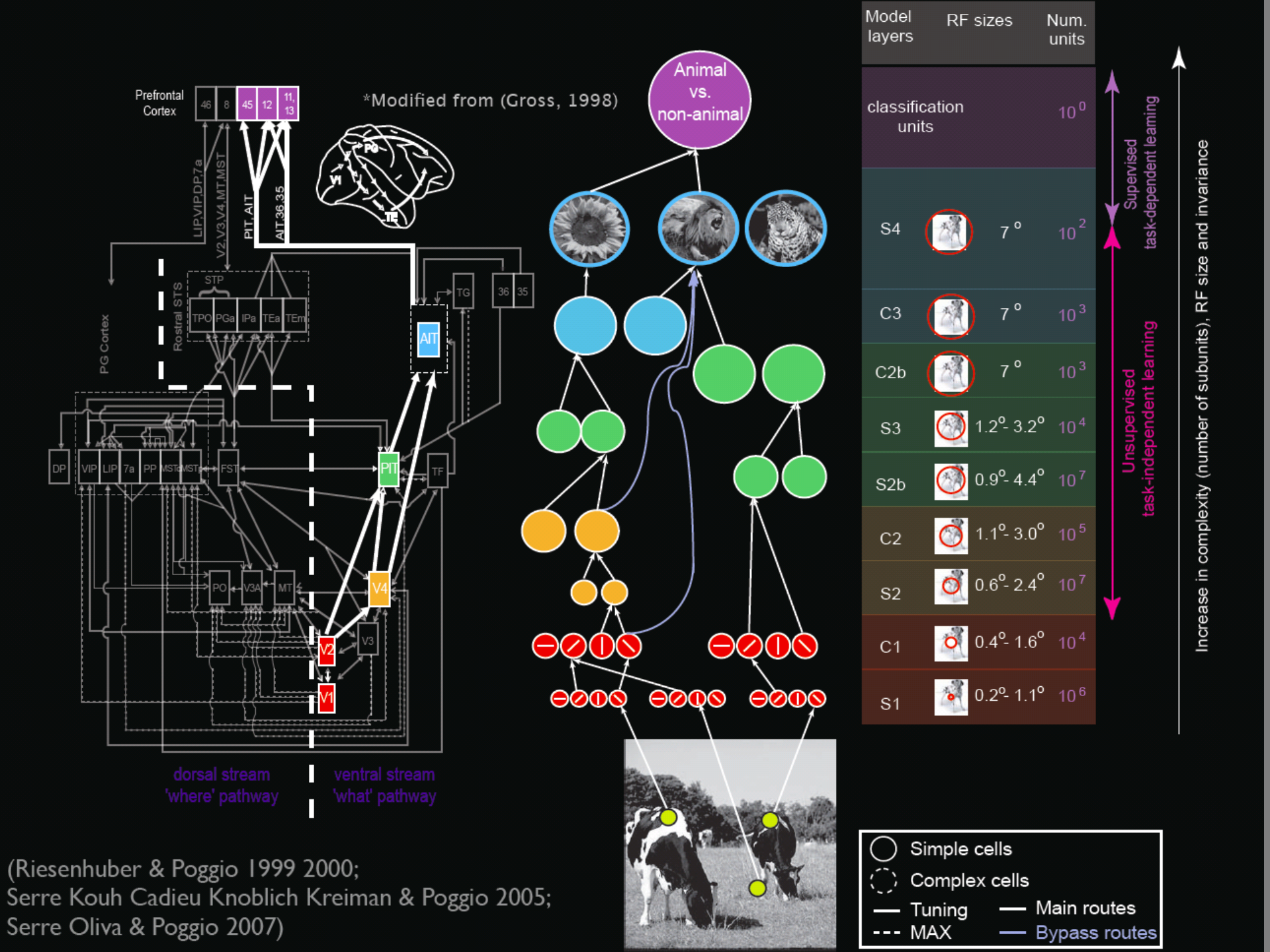
Examples of hammer:

How do we do it?





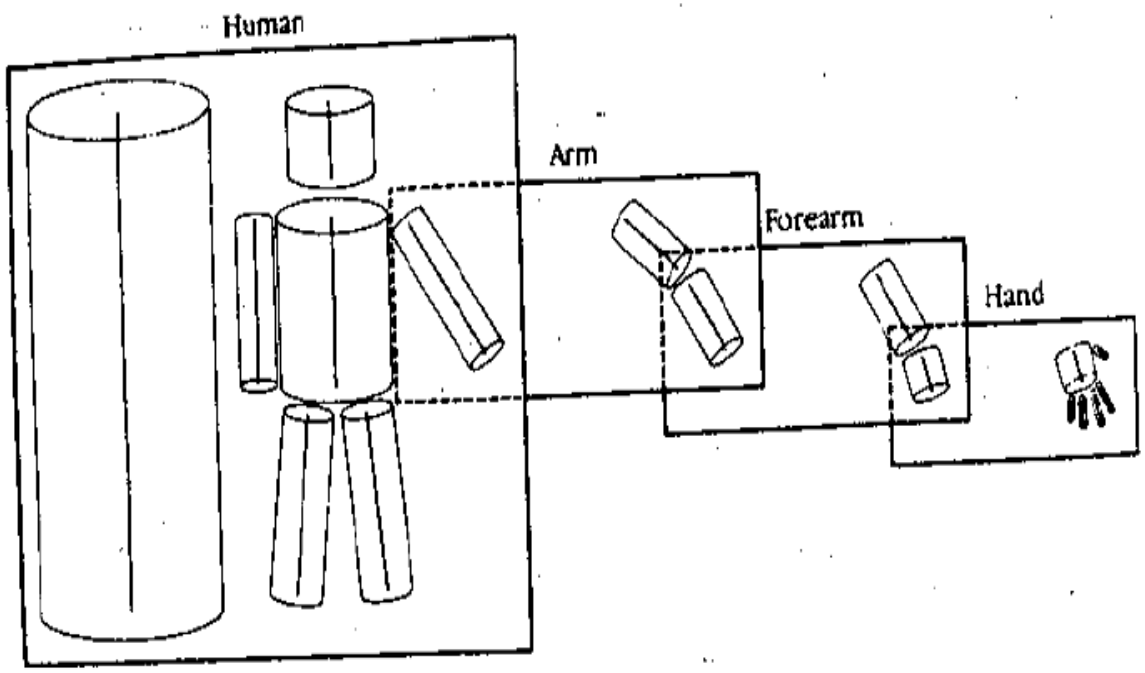
Felleman and Van Essen, 1991



(Riesenhuber & Poggio 1999 2000;
Serre Kouh Cadieu Knoblich Kreiman & Poggio 2005;
Serre Oliva & Poggio 2007)

Hierarchies in high-level vision

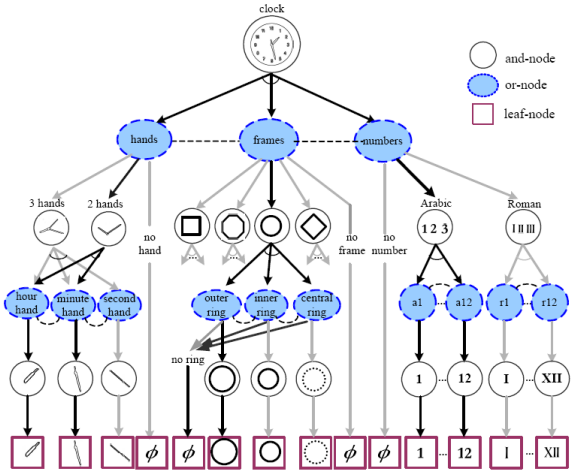
1980's pixels → edge → texton → motif → part → object

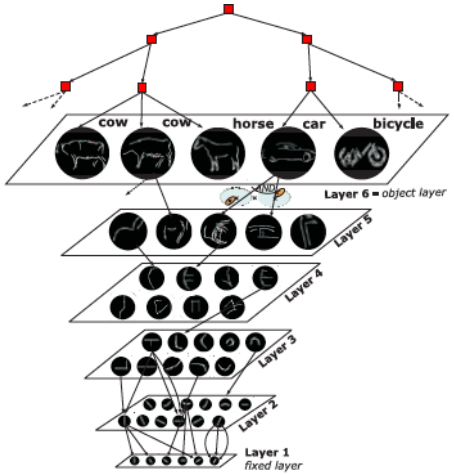


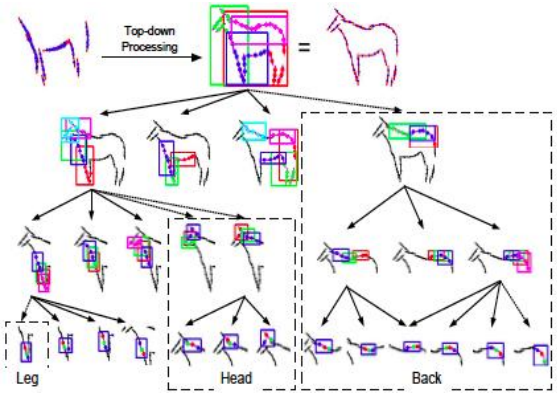
D. Marr and H. Nishihara, Representation and Recognition of the Spatial Organization of 3D Shapes, 1978
I. Biederman, Recognition-by-components: a theory of human image understanding,1987
S. Dickinson et. al., 3-D Shape Recovery Using Distributed Aspect Matching, 1992
S.C. Zhu and A. Yuille, FORMS: A Flexible Object Recognition and Modeling System, 1996.

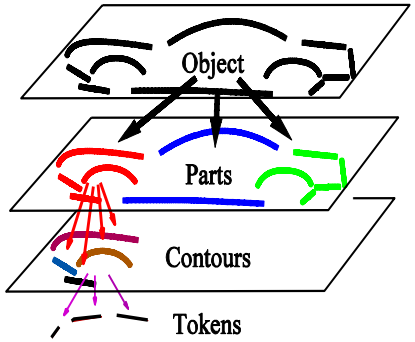
Hierarchies in high-level vision

2000-2010: probabilistic grammar models









S.C. Zhu and D. Mumford, `Quest for a Generic Grammar of Images', 2007
S. Filder and A. Leonardis, Towards Scalable Representations of Object Categories, CVPR 2007
L. Zhu and A. Yuille, Compositional Models, CVPR 2008-2010
I.Kokkinos and A. Yuille, Inference and Learning with Hierarchical Shape Models,CVPR 2009/IJCV 2011

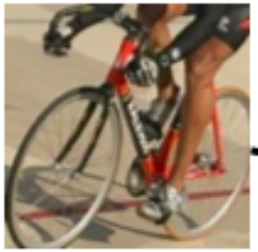
The real challenge: image features

$\mathbf{x}$ 

encoder Φ →

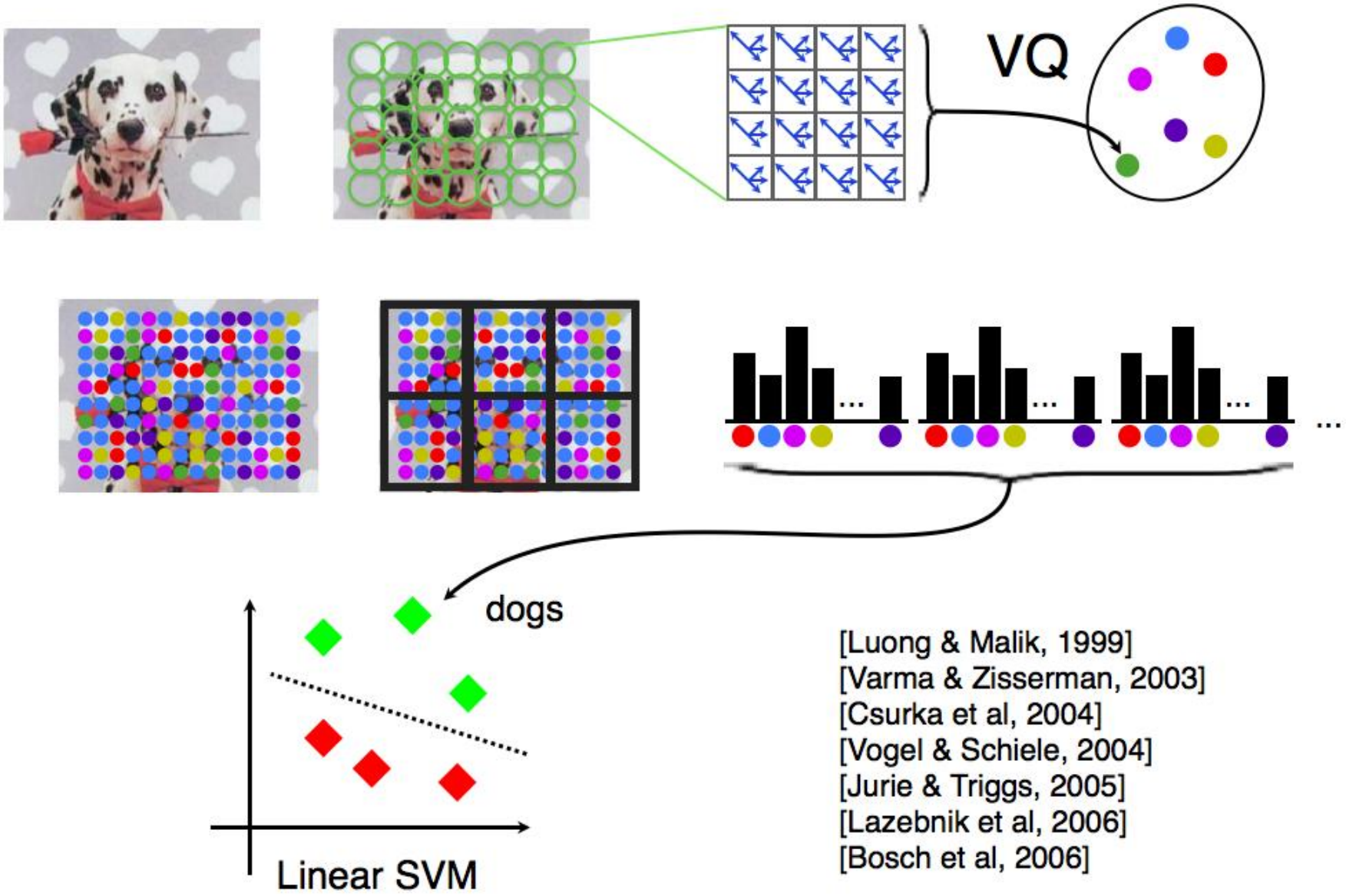
representation $\Phi(\mathbf{x}) \in \mathbb{R}^d$

Desirable feature properties



Φ is **invariant** to nuisance factors, **sensitive** to semantic variations

Image classification in a nutshell

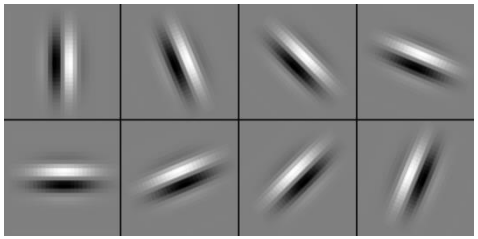


[Luong & Malik, 1999]
[Varma & Zisserman, 2003]
[Csurka et al, 2004]
[Vogel & Schiele, 2004]
[Jurie & Triggs, 2005]
[Lazebnik et al, 2006]
[Bosch et al, 2006]

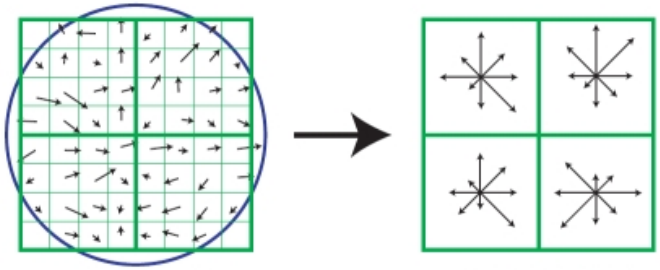
SIFT Descriptor

Image Pixels

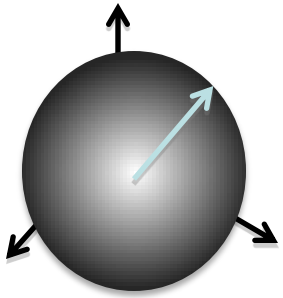
Apply Gabor filters



Spatial pool (Sum)



Normalize to unit length

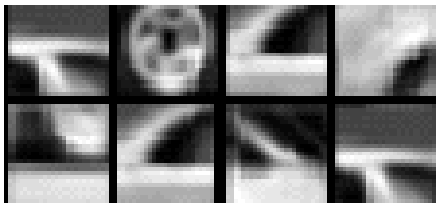


Feature Vector

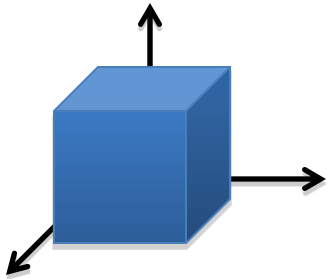
Spatial Pyramid Matching

SIFT Features

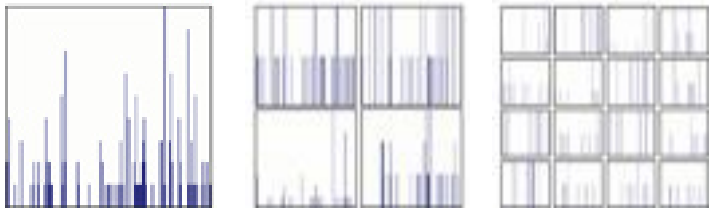
Filter with Visual Words



Max



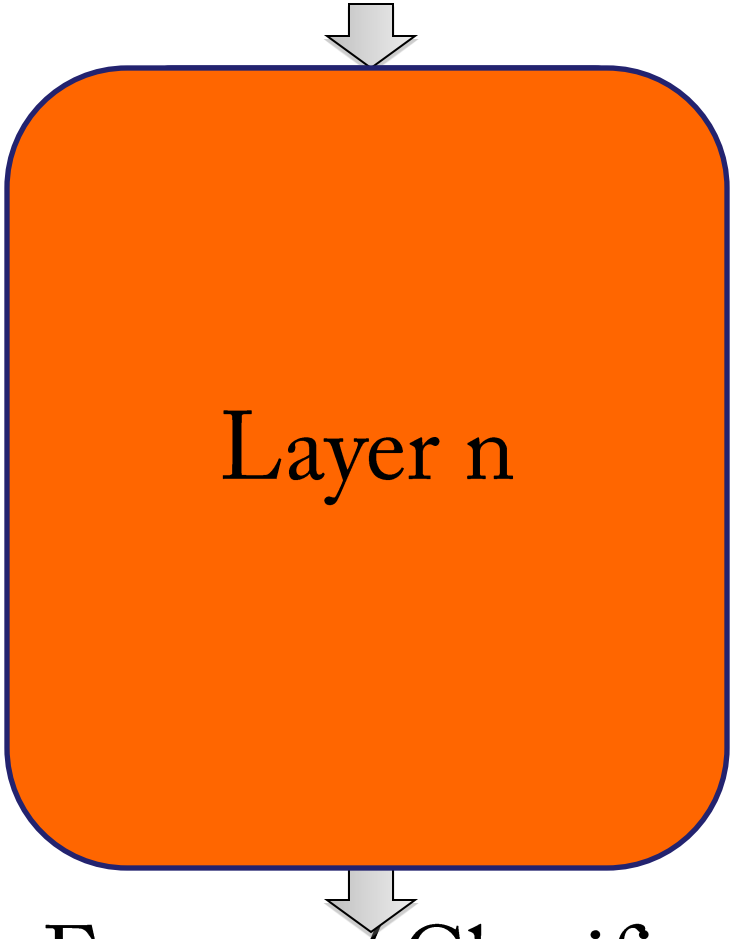
Multi-scale spatial pool (Sum)



Classifier

Single Layer Architecture

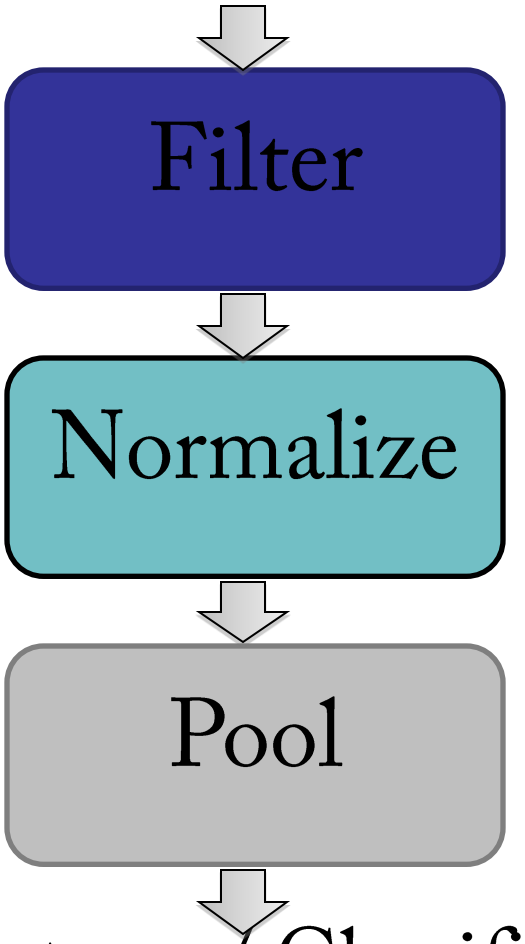
Input: Image Pixels / Features



Output: Features / Classifier

Single Layer Architecture

Input: Image Pixels / Features



Output: Features / Classifier

Deep Learning and Computer Vision

1980's pixels → edge → texton → motif → part → object

2000-2010

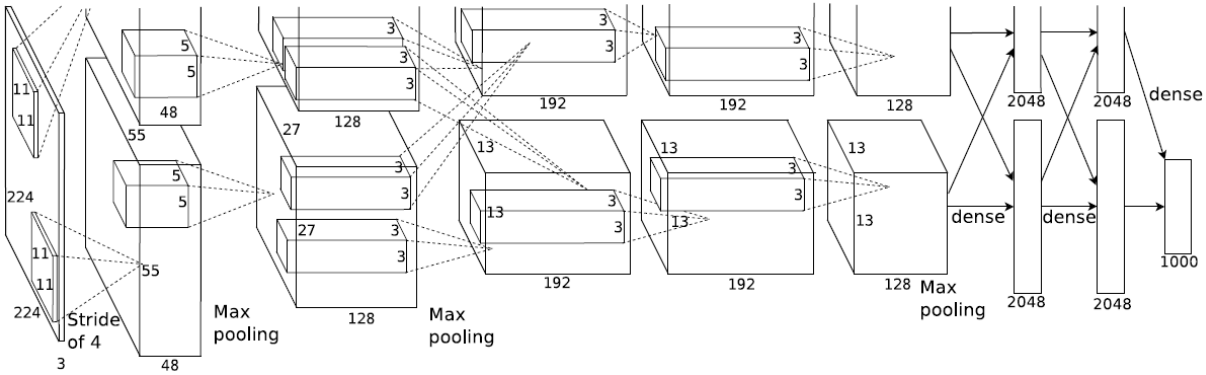
SIFT/HOG → K-Means/ pooling → classifier → "car"

fixed unsupervised supervised

2010+

$x \rightarrow max(0, W^1 x) \rightarrow h^1 \rightarrow max(0, W^2 h^1) \rightarrow h^2 \rightarrow W^3 h^2 \rightarrow o$

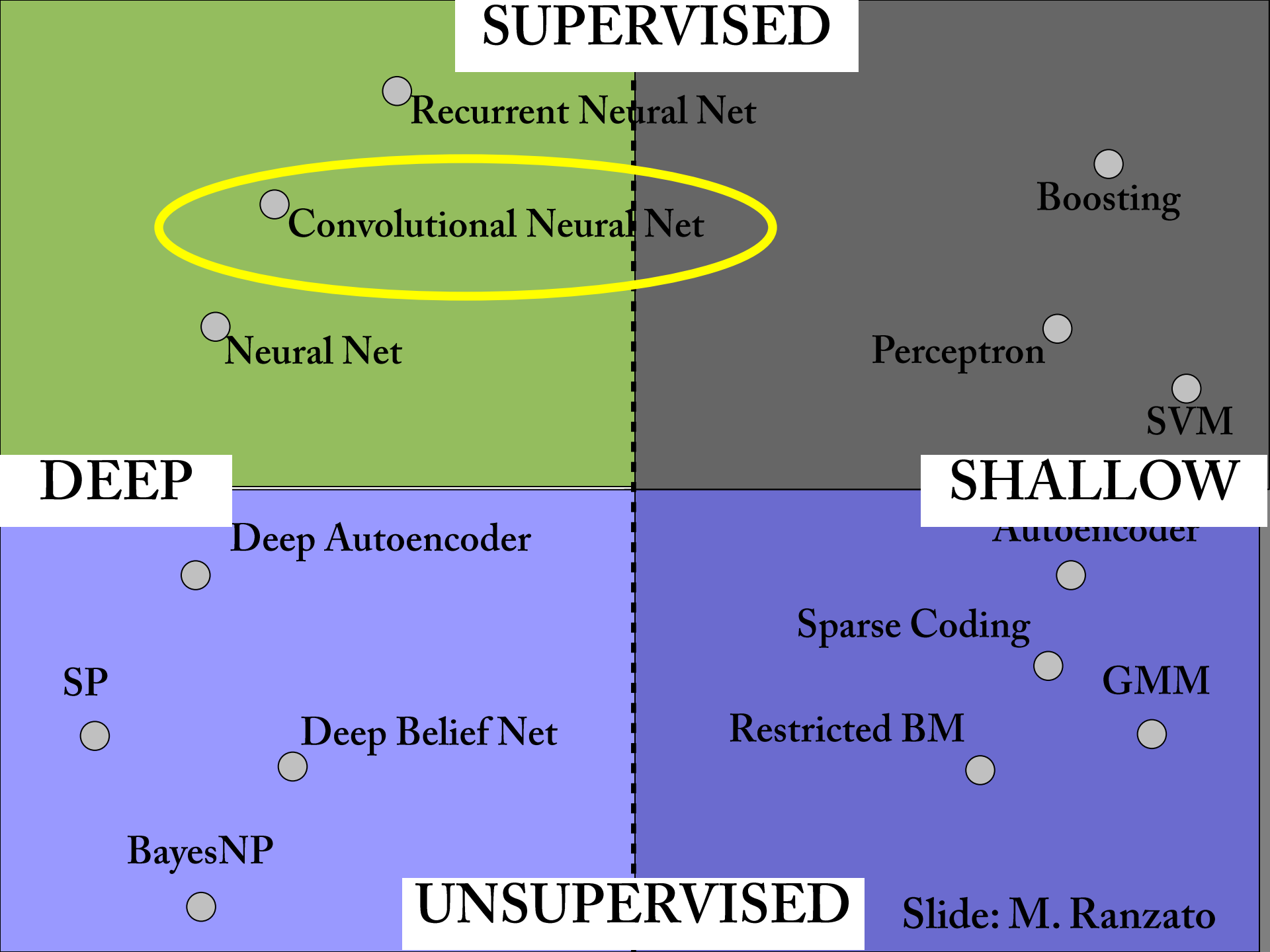
Breakthrough: Imagenet 2012



A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*13

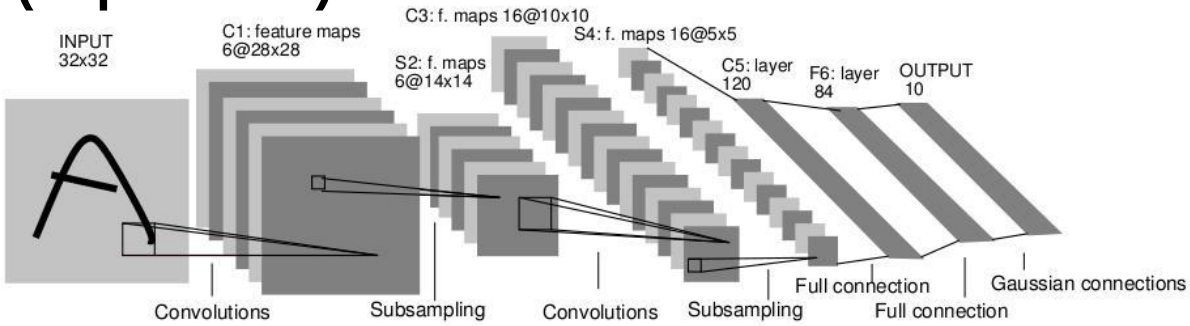
A taxonomy of methods

- We ill see some of them in detail in this class


SUPERVISED
Recurrent Neural Net
Convolutional Neural Net
Neural Net
Boosting
Perceptron
SVM
DEEP
SHALLOW
Deep Autoencoder
Sparse Coding
SP
GMM
Deep Belief Net
Restricted BM
BayesNP
UNSUPERVISED
Slide: M. Ranzato

Convolutional Networks

Discriminative (supervised)



Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541-551, Winter 1989

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998.

Generative (unsupervised)

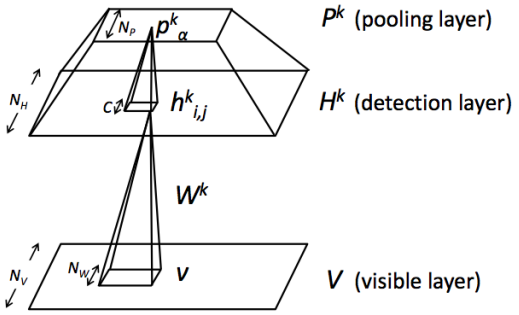


Figure 1. Convolutional RBM with probabilistic max-pooling. For simplicity, only group k of the detection layer and the pooing layer are shown. The basic CRBM corresponds to a simplified structure with only visible layer and detection (hidden) layer. See text for details.

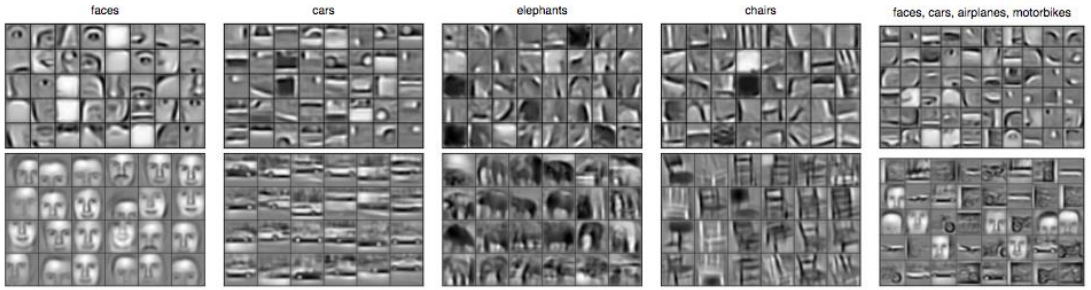
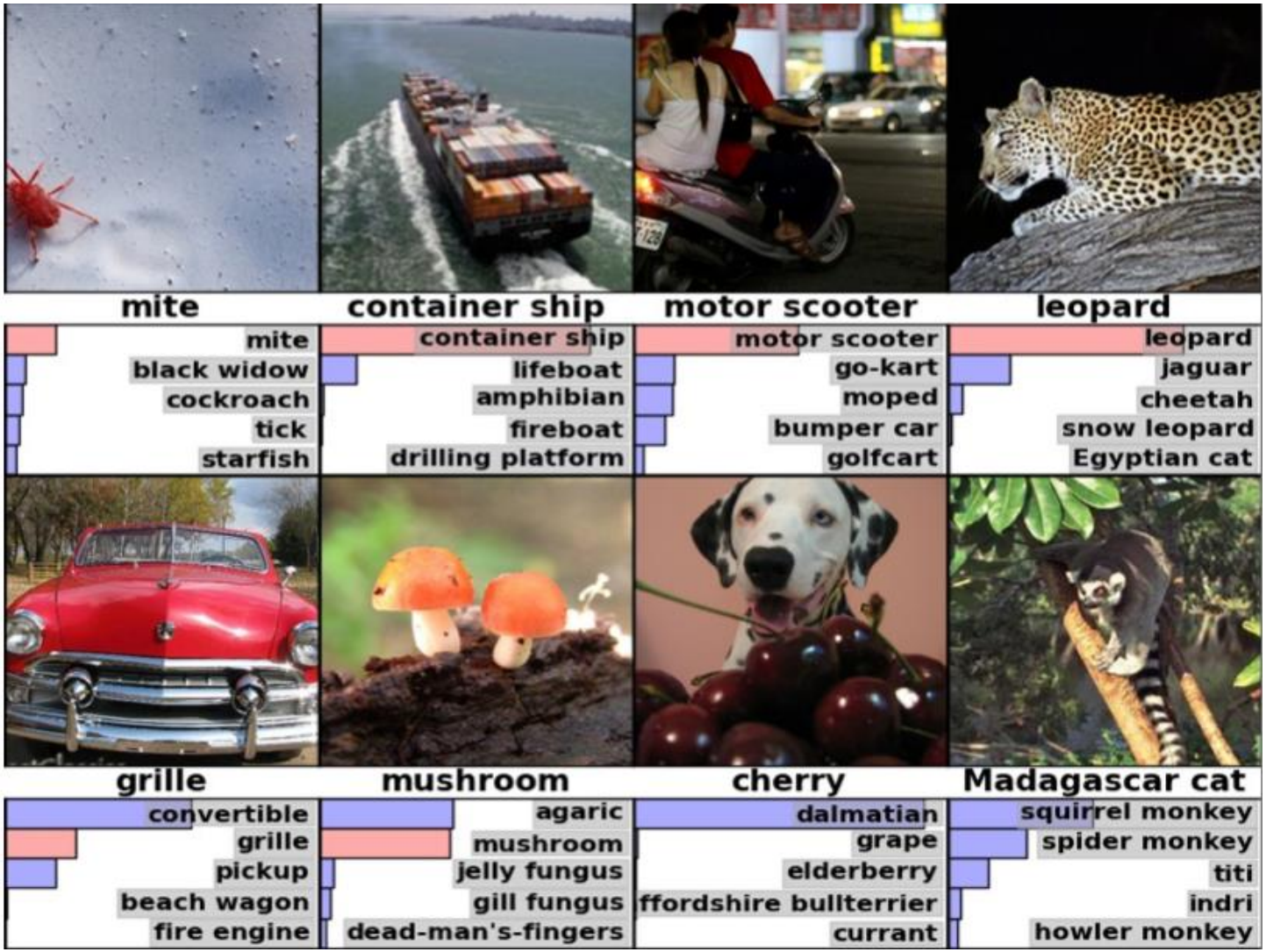


Figure 3. Columns 1-4: the second layer bases (top) and the third layer bases (bottom) learned from specific object categories. Column 5: the second layer bases (top) and the third layer bases (bottom) learned from a mixture of four object categories (faces, cars, airplanes, motorbikes).

Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations, H. Lee, R. Grosse, R. Ranganath, A. Y. Ng, ICML 2010

Imagenet top-5 error rate: 36%-> 18% (‘12) -> 6% (‘14)



A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*13

Demo from http://www.clarifai.com/

Predicted Tags:

newborn
baby
infant
bed
terrain
root vegetable
bedroom
family
sleep
innocence

Stats:

Size: 22.20 KB
Time: 70 ms

Similar Images:

Demo from http://www.clarifai.com/

Predicted Tags:

sculpture
ham
heavy
iron
history
painting
steel
war
face

Stats:

Size: 42.80 KB
Time: 55 ms

The black shape to the right of the fish looks like an oncoming airplane, and there is a trail of smoke in the right part of the sky. These may be allusions to the "mechanical terror of the war experience" which led to Ernst writing, "On the 1st of August 1914 Max Ernst died. He was resurrected on the 11 November 1918 as a young man who aspired to find the myths of his time." Celebes, then, seems to represent the myth of destruction

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

Demo from http://www.clarifai.com/

Predicted Tags:

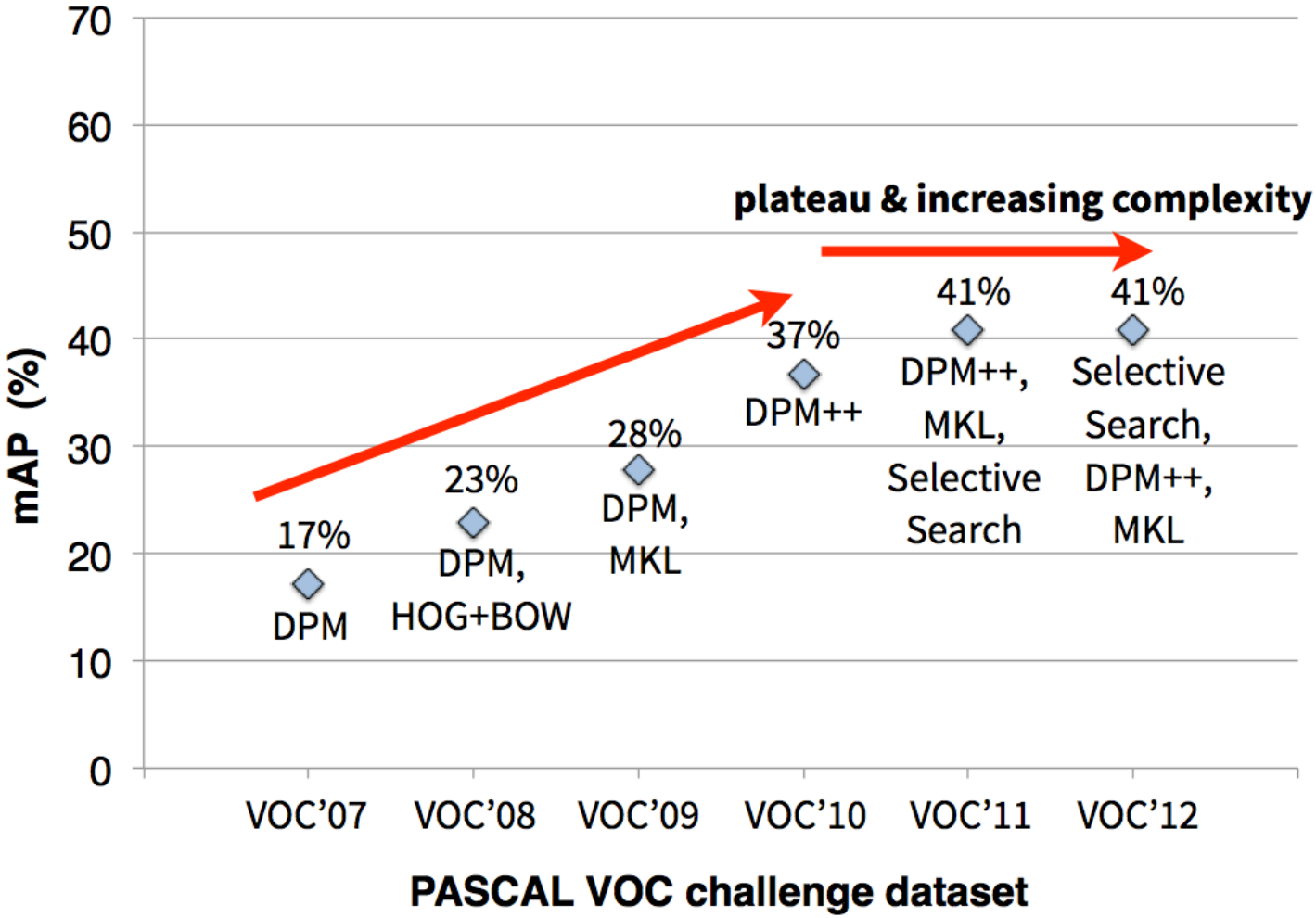
zoo
courage
owl
wildlife
bird
animal
forest
ape
tree

Stats:

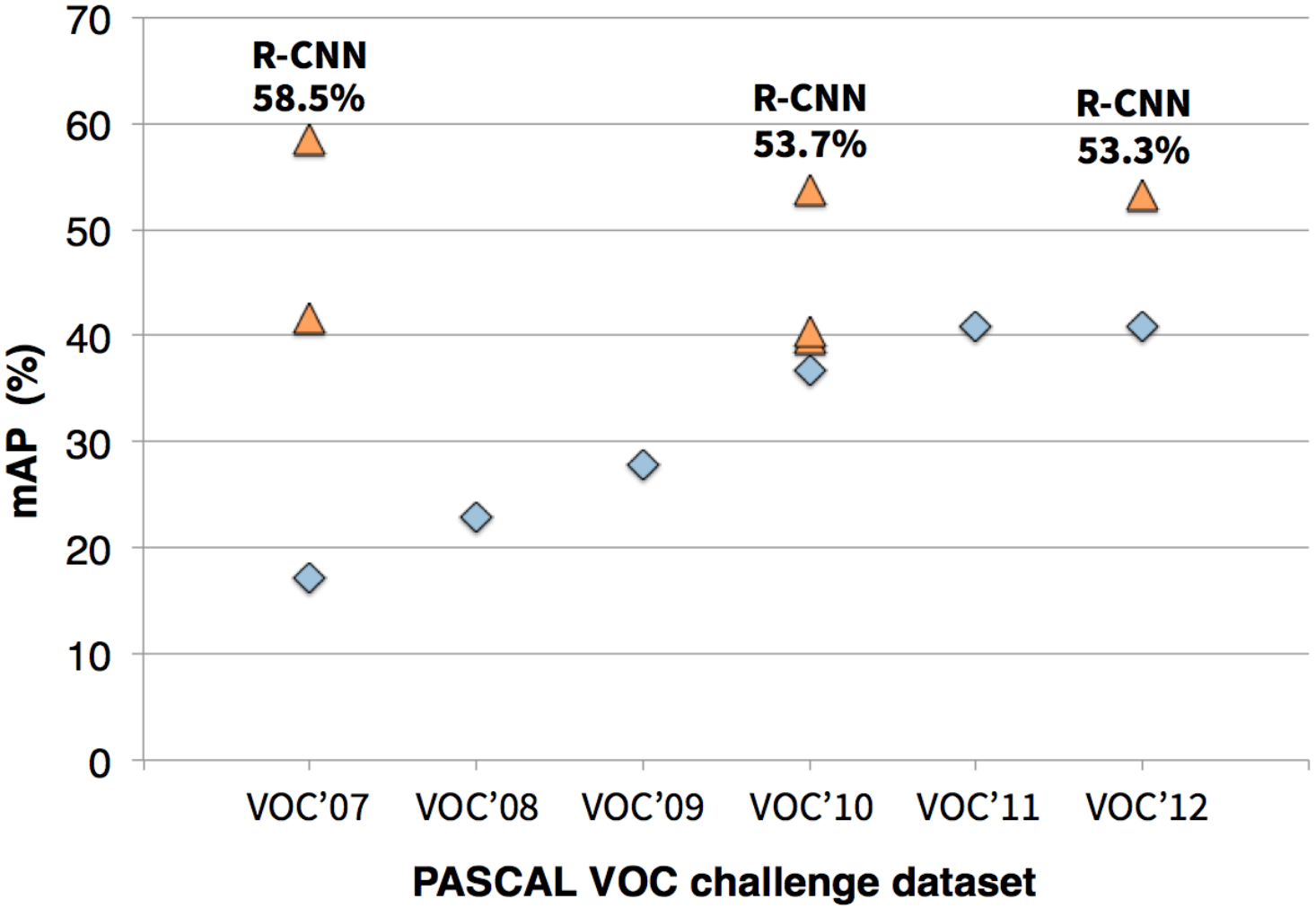
Size: 206.04 KB
Time: 78 ms

Similar Images:

Object recognition 2010-2012: saturation



Object recognition, 2013

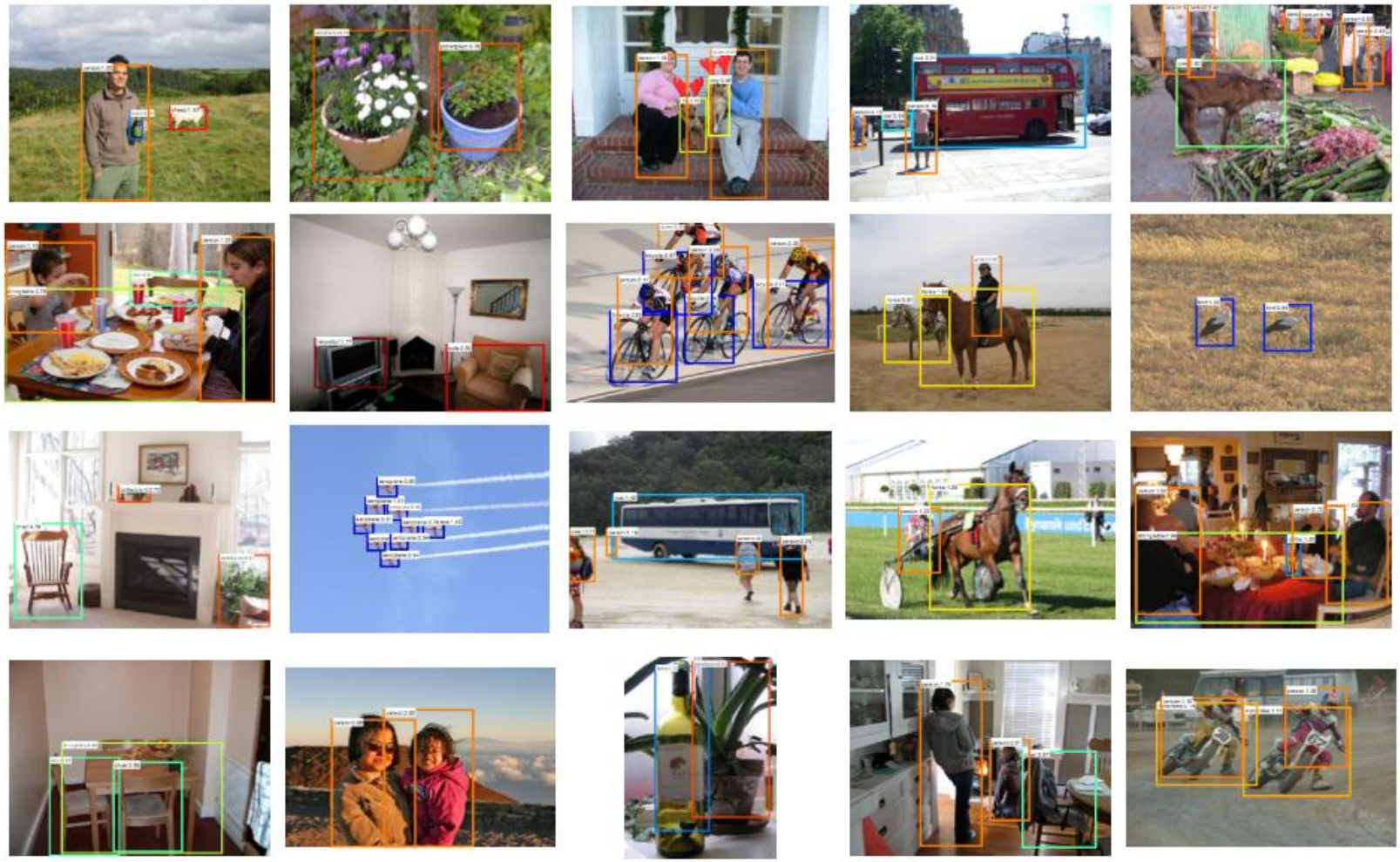


A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*13

P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. ICLR 14.

R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR* 14

Object recognition, 2014



A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet classification with deep convolutional neural networks. *NIPS*13

P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. ICLR 14.

R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *CVPR* 14

Deep Learning and AI

VISION

pixels → edge → texton → motif → part → object

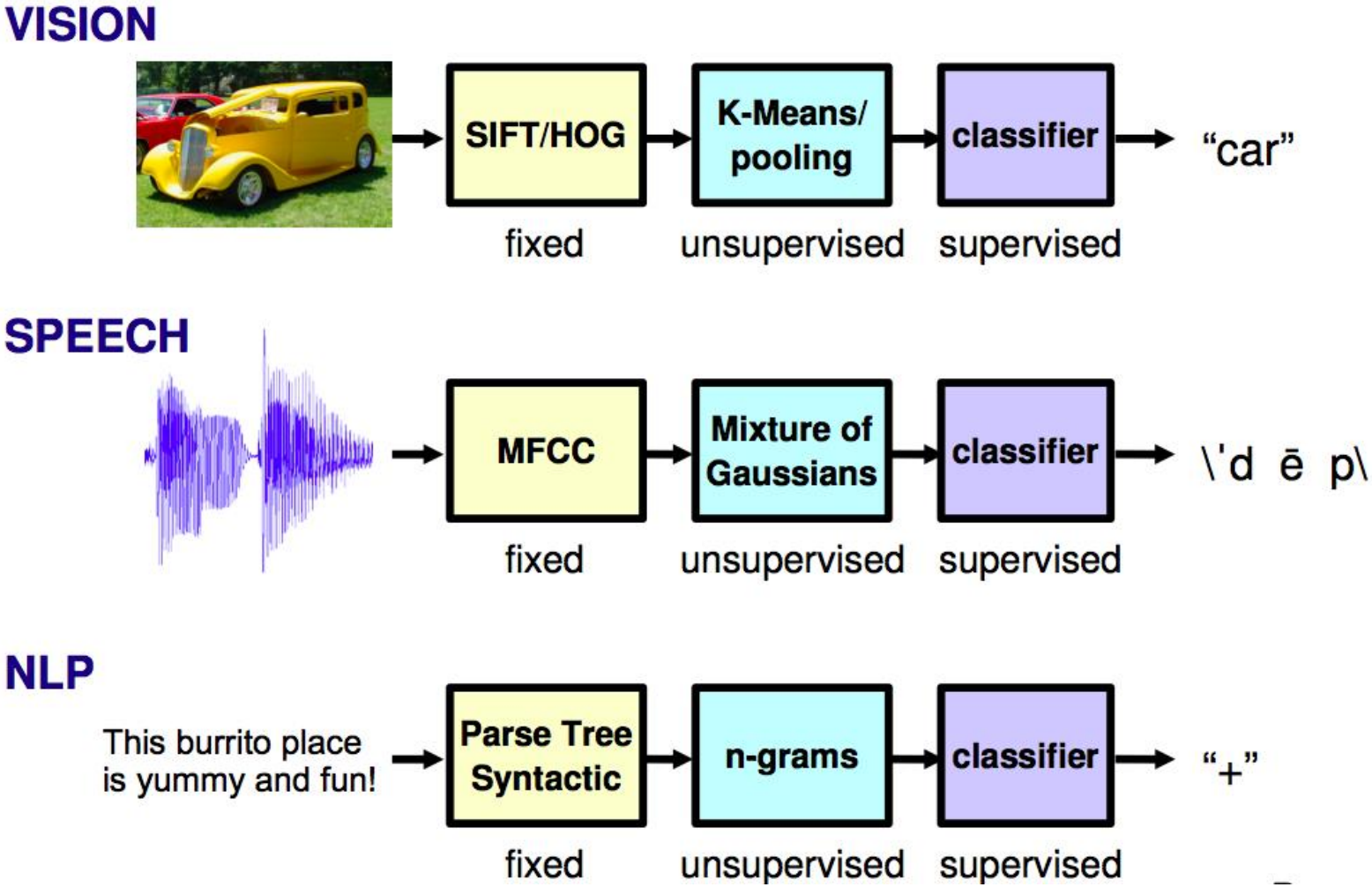
SPEECH

sample → spectral band → formant → motif → phone → word

NLP

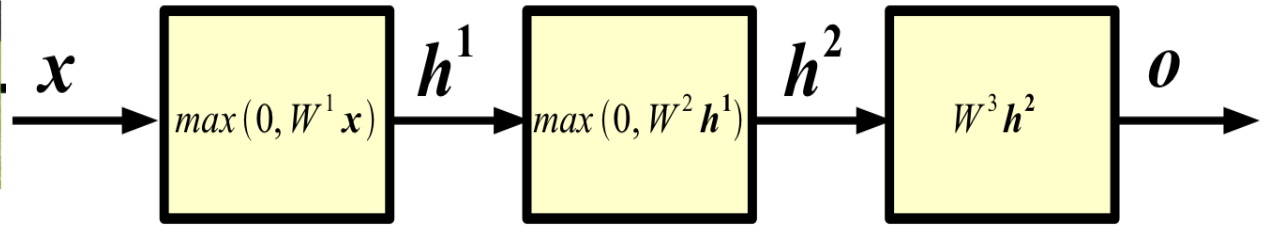
character → word → NP/VP/.. → clause → sentence → story

Deep Learning and AI

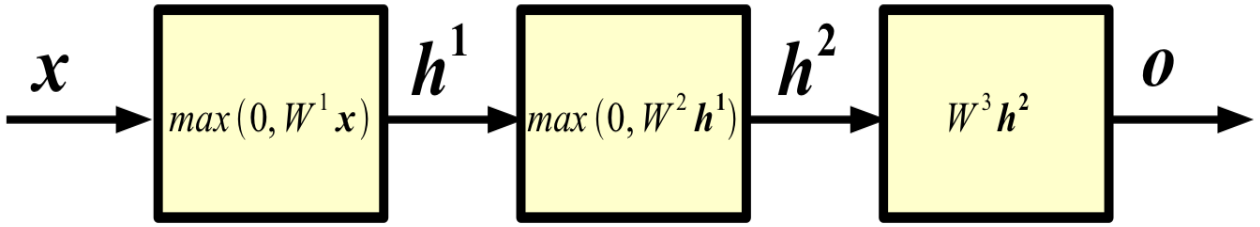


Deep Learning and AI

VISION

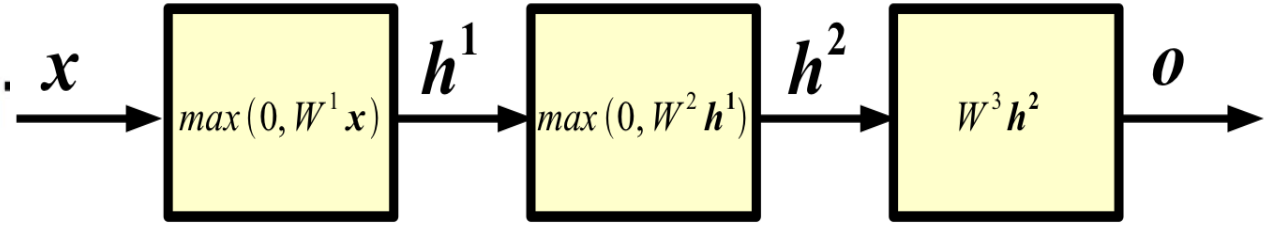


SPEECH



NLP

This burrito place is yummy and fun!



'Deep Learning reads Wikipedia and discovers the meaning of life'
Geoff Hinton

Deep Learning in the media

WIRED

Researcher Dreams Up Machines That Learn Without Humans
06.27.13

The New York Times

Scientists See Promise in Deep-Learning Programs
John Markoff
November 23, 2012

THE GLOBE AND MAIL

Google taps U of T professor to teach context to computers
03.11.13

19

WIRED

The Man Behind the Google Brain: Andrew Ng and the Quest for the New AI

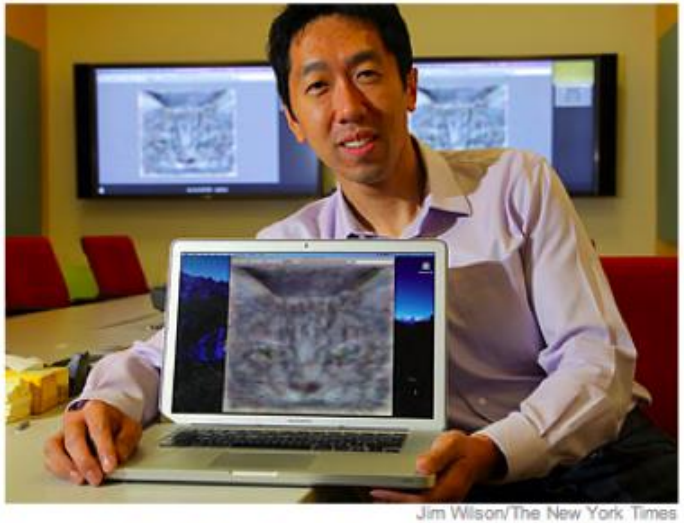
BY DANIELA HERNANDEZ 05.07.13 6:30 AM

The New York Tim

Monday, June 25, 2012 Last Update: 11:50 PM ET

DIGITAL SUBSCRIPTION: 4 WEEKS

ING DIRECT Follow Us



Jim Wilson/The New York Times

Despite Itself, a Simulated Brain Seeks Cats

By JOHN MARKOFF 12 minutes ago

A Google research team, led by Andrew Y. Ng, above, and Jeff Dean, created a neural network of 16,000 processors that reflected human obsession with Internet felines.

3 Lectures:

- < Deep Learning> : Origins
  - Initial ideas: What worked and what didnt
- < Deep Learning> : Modern Deep Architectures
  - What are the factors that led to current success
- < Deep Learning> : Computer Vision Applications
  - CNN’s
  - RCNN’s
  - Object Classification,
  - Action recognition
  - Medical Image Analysis

Projects

- MatConvNet: http://www.vlfeat.org/matconvnet/
- MATLAB toolbox implementing *Convolutional Neural Networks* (CNNs) for computer vision applications. It is simple, efficient, and can run and learn state-of-the-art CNNs. Several example CNNs are included to classify and encode images.
- Other popular packages?
  - CAFFE
  - VGG
  - Theano
  - Overfeat
  - Torch

Project Ideas

With MatConv Net:

1. Hu, Guosheng, et al. "When Face Recognition Meets with Deep Learning: an Evaluation of Convolutional Neural Networks for Face Recognition." *arXiv preprint arXiv:1504.02351* (2015).http://arxiv.org/abs/1504.02351

Face recognition with MatConvNet. The network is really small. You can train the small version of CNN in the paper in a day, with CPU

2. Srivastava, Nitish, et al. "Dropout: A simple way to prevent neural networks from overfitting." *The Journal of Machine Learning Research* 15.1 (2014): 1929-1958.

http://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf

Evaluating the dropout method on MNIST dataset. Since the dataset is small, they can train the CNN in a day, with CPU.

3. Oquab, Maxime, et al. "Learning and transferring mid-level image representations using convolutional neural networks." *Computer Vision and Pattern Recognition (CVPR), 2014*.

http://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Oquab_Learning_and_Transferring_2014_CVPR_paper.pdf

Image classification on the Pascal VOC dataset. Use the CNN learnt on ImageNet to extract features. On top of the features, train a 2 layer neural network to classify the images.

The pretrained model is available in MatConvNet, so no need to train a CNN. If Pascal VOC is too large try Caltech 101.

More Project Ideas

1. Test out the fast RCNN paper: http://arxiv.org/pdf/1504.08083v1.pdf

caffe with python/matlab interfaces: https://github.com/rbgirshick/fast-rcnn

What should you do:

find a new dataset, for example, SUN dataset, fine tune the parameters using the framework the author provided, then run on the test set to see the results.

2. Go to Kaggle.com and picj a challenge! That way you can compare your performance