Introduction: Convolutional Neural Networks for Visual Recognition

boris .ginzburg@intel.com

Acknowledgments

This presentation is heavily based on:

- http://cs.nyu.edu/~fergus/pmwiki/pmwiki.php
- http://deeplearning.net/reading-list/tutorials/
- http://deeplearning.net/tutorial/lenet.html
- http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial

... and many other

Agenda

1. Course overview
2. Introduction to Deep Learning
   - Classical Computer Vision vs. Deep learning
3. Introduction to Convolutional Networks
   - Basic CNN Architecture
   - Large Scale Image Classifications
   - How deep should be Conv Nets?
   - Detection and Other Visual Apps

Course overview

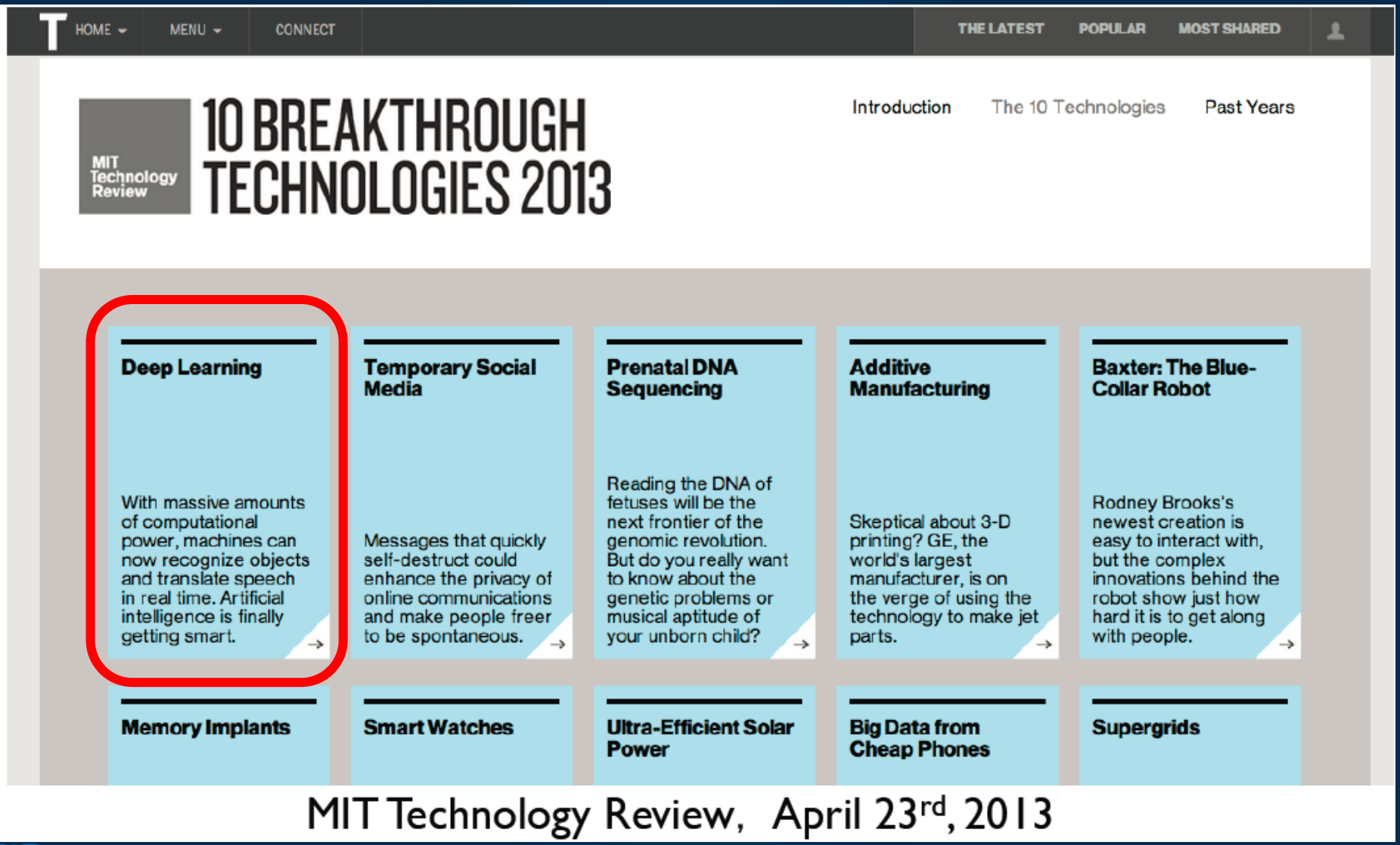
1. Introduction
   - Intro to Deep Learning
   - Caffe: Getting started
   - CNN: network topology, layers definition
2. CNN Training
   - Backward propagation
   - Optimization for Deep Learning: SGD : monentum, rate adaptation, Adagrad, SGD with Line Search, CGD
   - “Regularization” (Dropout , Maxout)

Course overview

3. Localization and Detection
   - Overfeat
   - R-CNN (Regions with CNN)
4. CPU / GPU performance optimization
   - CUDA
   - Vtune, OpenMP, and BLAS/MKL

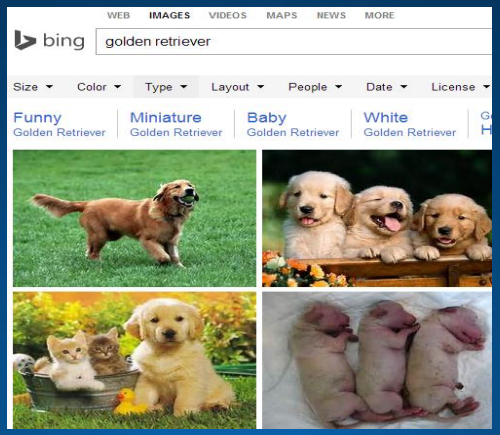
Introduction to Deep Learning

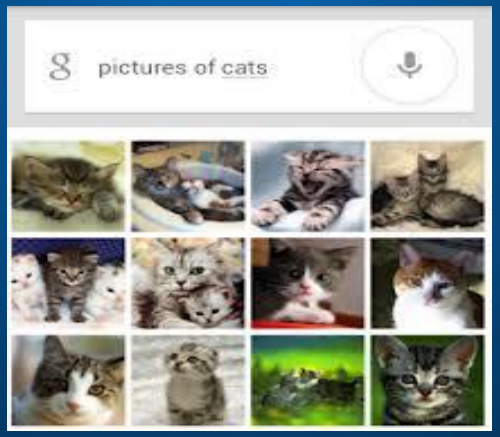
Buzz...



MIT Technology Review, April 23rd, 2013

Deep Learning – from Research to Technology





Deep Learning - breakthrough in visual and speech recognition

Classical Computer Vision Pipeline

Classical Computer Vision Pipeline.

CV experts

1. Select / develop features: SURF, HoG, SIFT, RIFT, …
2. Add on top of this Machine Learning for multi-class recognition and train classifier

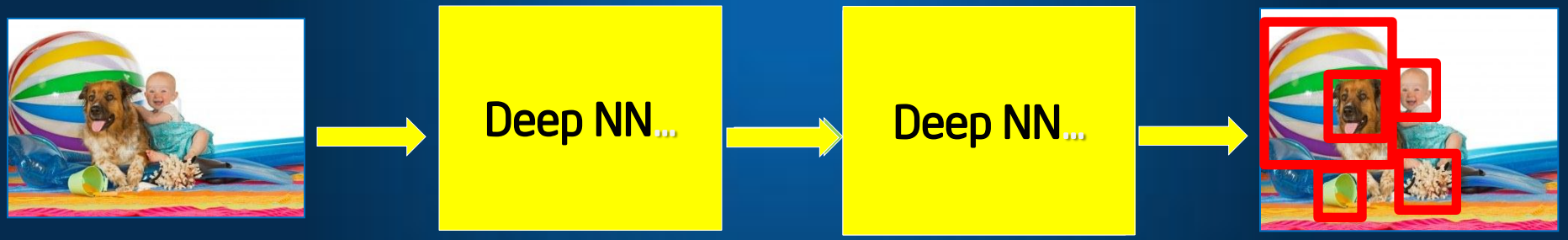
Classical CV feature definition is domain-specific and time-consuming

Deep Learning –based Vision Pipeline.

Deep Learning:

- Build features automatically based on training data
- Combine feature extraction and classification

DL experts: define NN topology and train NN



Deep Learning promise:
train good feature automatically,
same method for different domain

Computer Vision +Deep Learning + Machine Learning

We want to combine Deep Learning + CV + ML

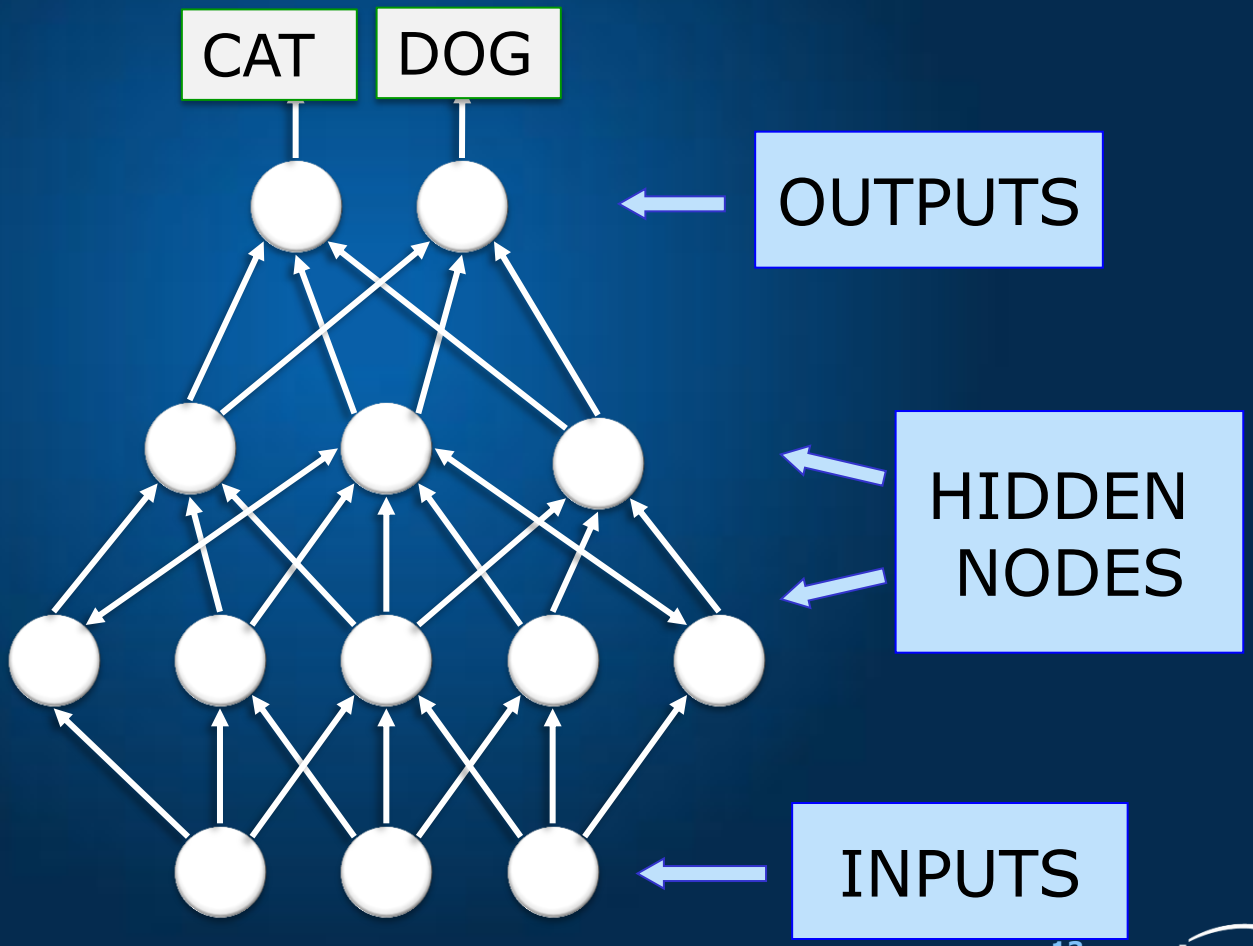
- Combine pre-defined features with learned features;
- Use best ML methods for multi-class recognition

CV+DL+ML experts needed to build the best-in-class

Combine best of Computer Vision
Deep Learning and Machine Learning

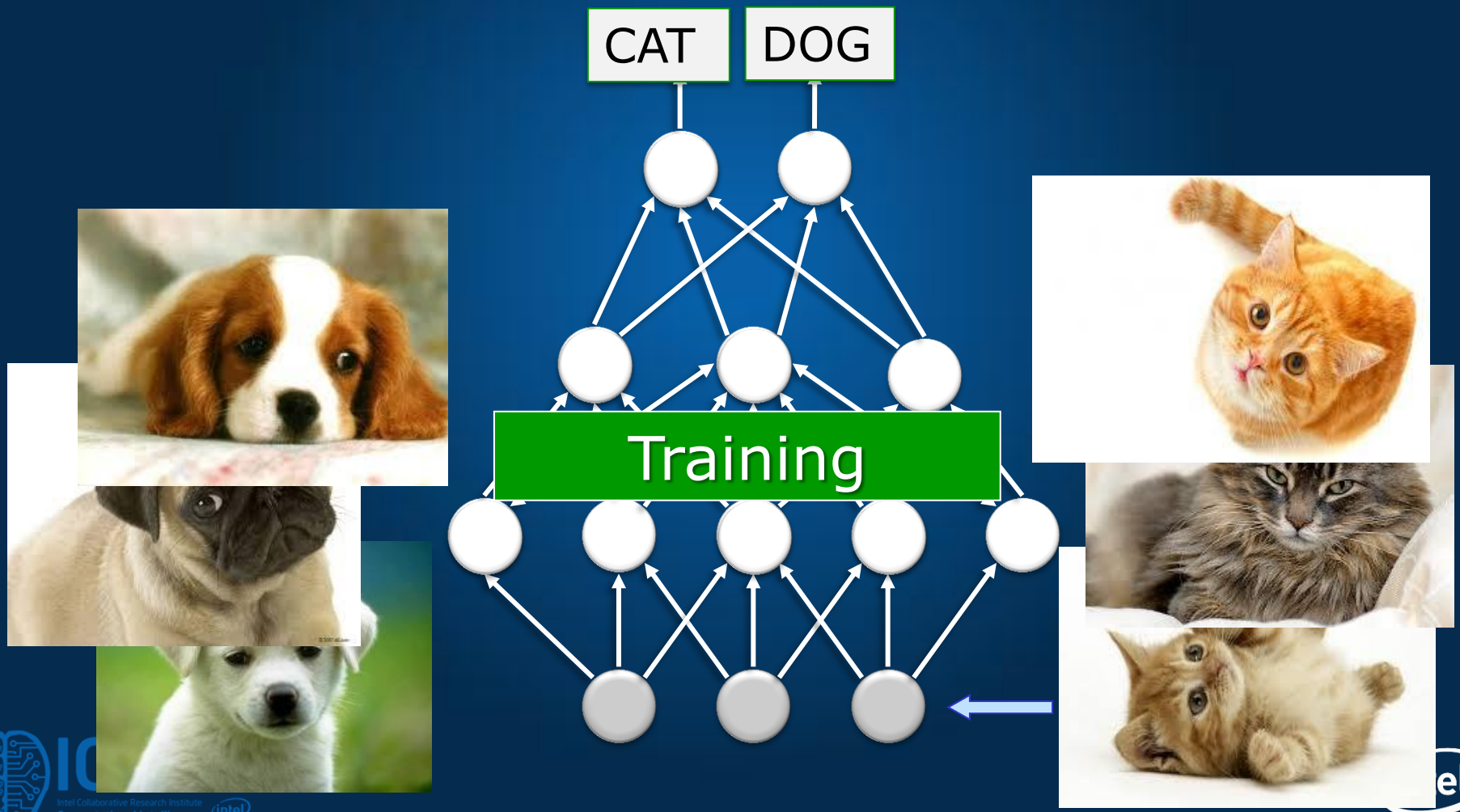
Deep Learning Basics

Deep Learning – is a set of machine learning algorithms based on multi-layer networks



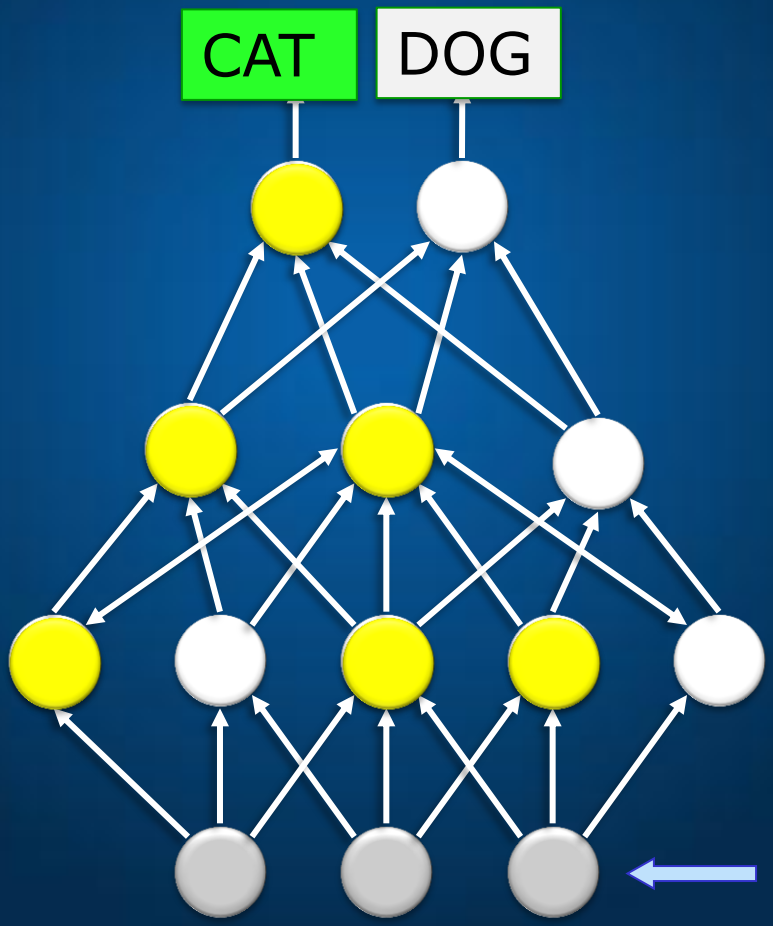
Deep Learning Basics

Deep Learning – is a set of machine learning algorithms based on multi-layer networks



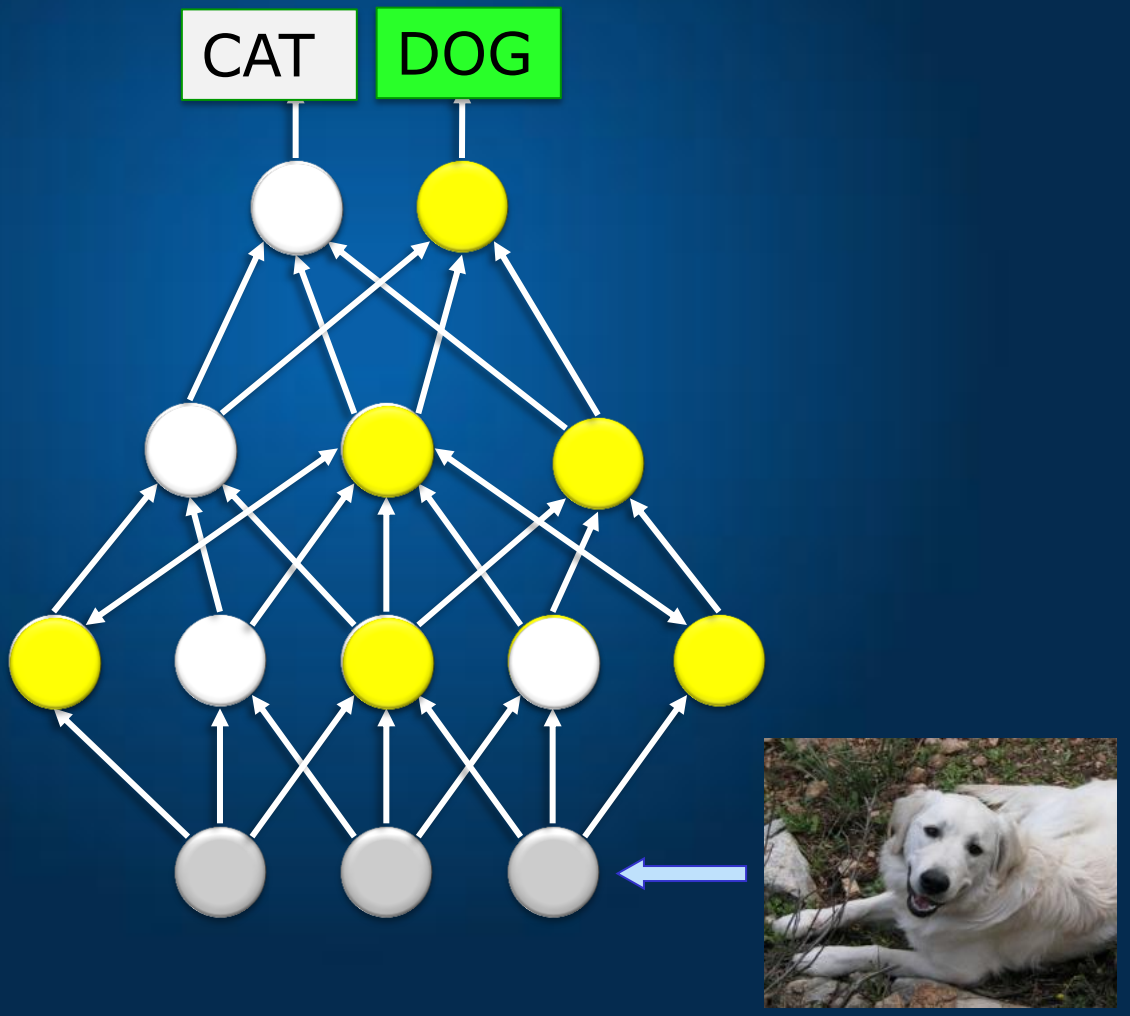
Deep Learning Basics

Deep Learning – is a set of machine learning algorithms based on multi-layer networks



Deep Learning Basics

Deep Learning – is a set of machine learning algorithms based on multi-layer networks



Deep Learning Taxonomy

Supervised:

- Convolutional NN (LeCun)
- Recurrent Neural nets (Schmidhuber)

Unsupervised

- Deep Belief Nets / Stacked RBMs (Hinton)
- Stacked denoising autoencoders (Bengio)
- Sparse AutoEncoders (LeCun, A. Ng,)

Convolutional Networks

Convolutional NN

Convolutional Neural Networks is extension of traditional Multi-layer Perceptron, based on 3 ideas:

1. Local receive fields
2. Shared weights
3. Spatial / temporal sub-sampling

See LeCun paper (1998) on text recognition:

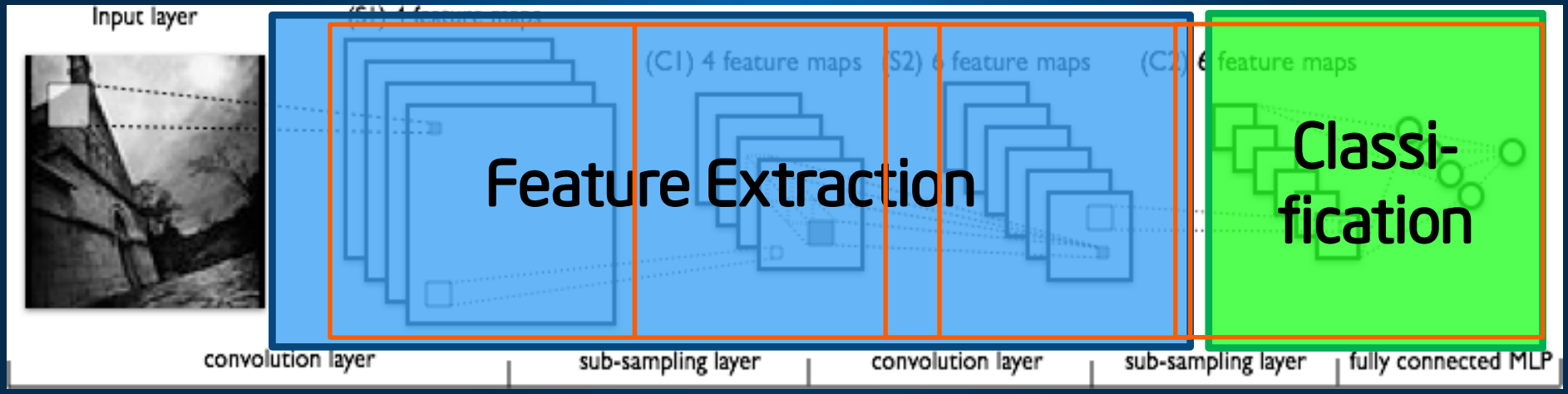
http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf

What is Convolutional NN ?

CNN - multi-layer NN architecture

- Convolutional + Non-Linear Layer
- Sub-sampling Layer
- Convolutional +Non-L inear Layer
- Fully connected layers

• Supervised



What is Convolutional NN ?

CNN story: 1996 - MNIST

Lenet-5 (1996) : core of CNR check reading system, used by US banks.

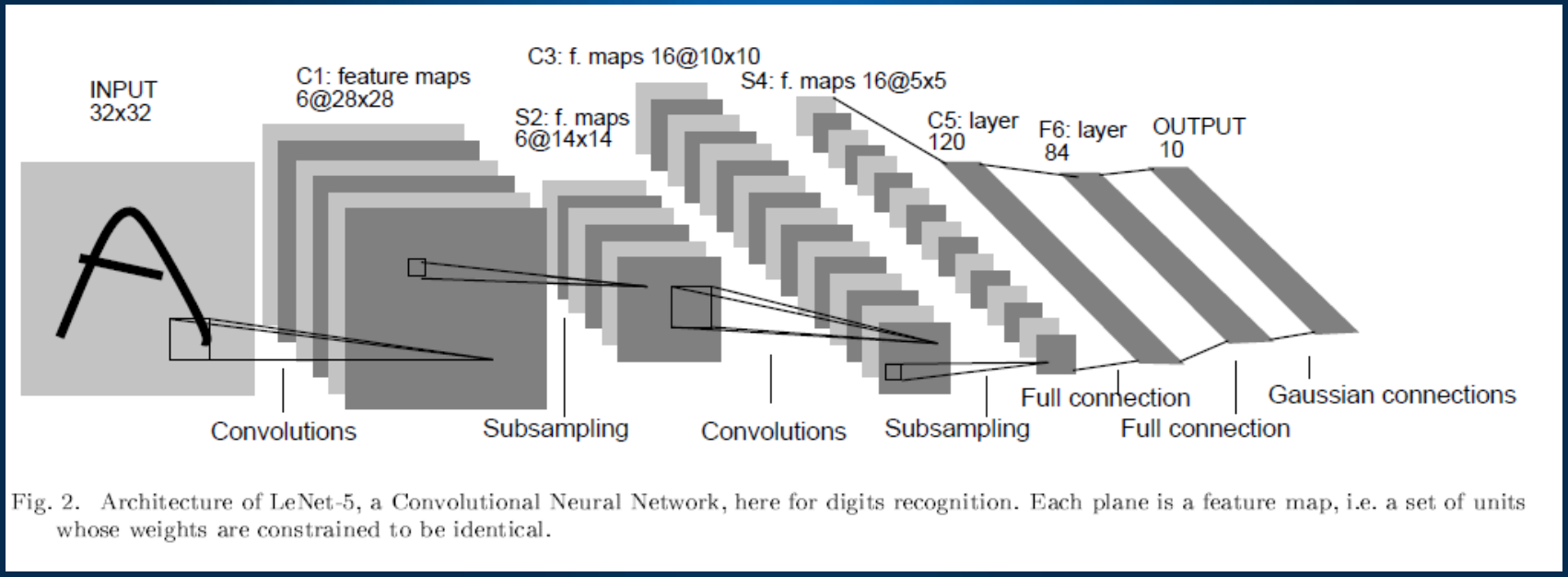
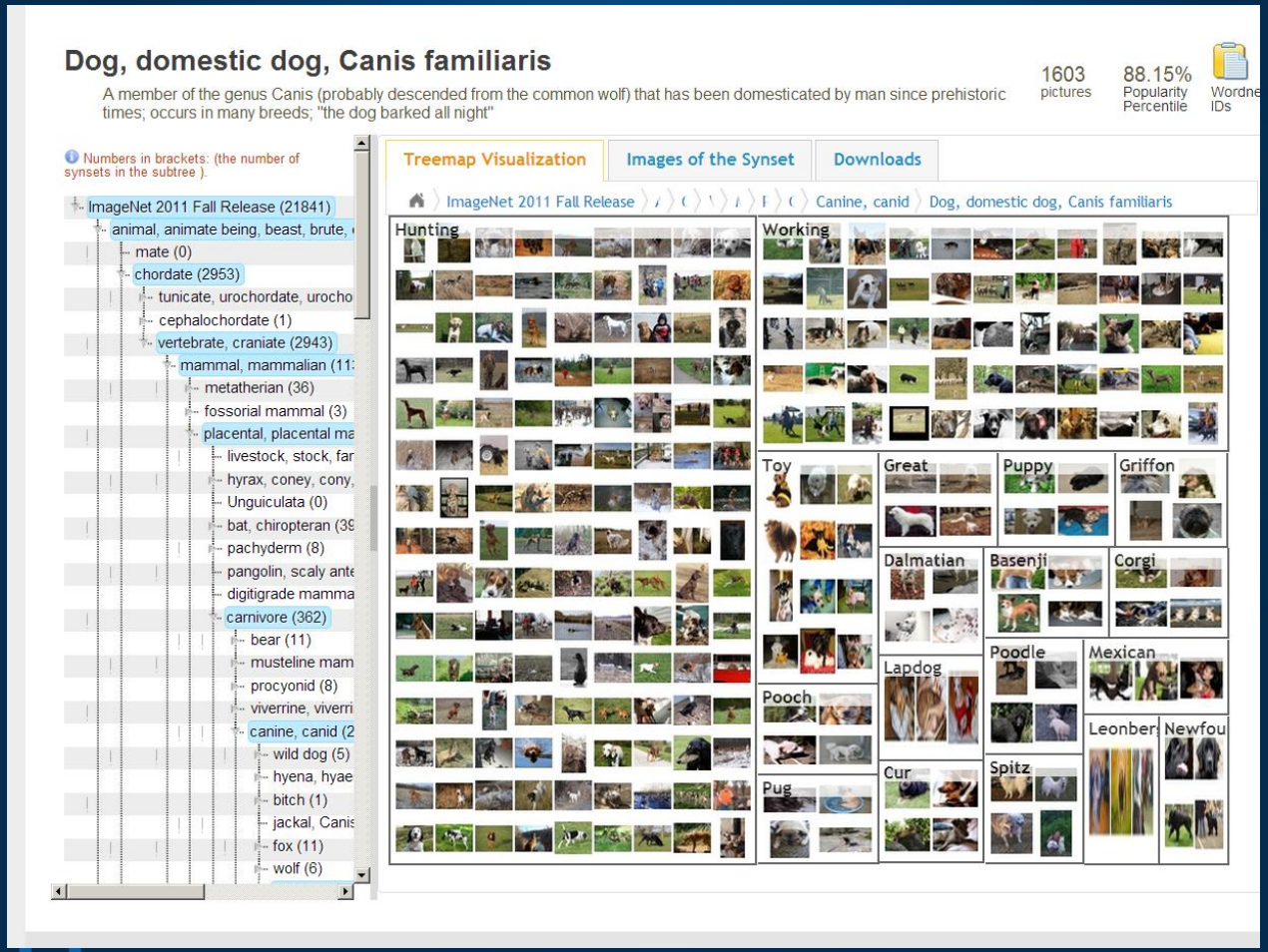


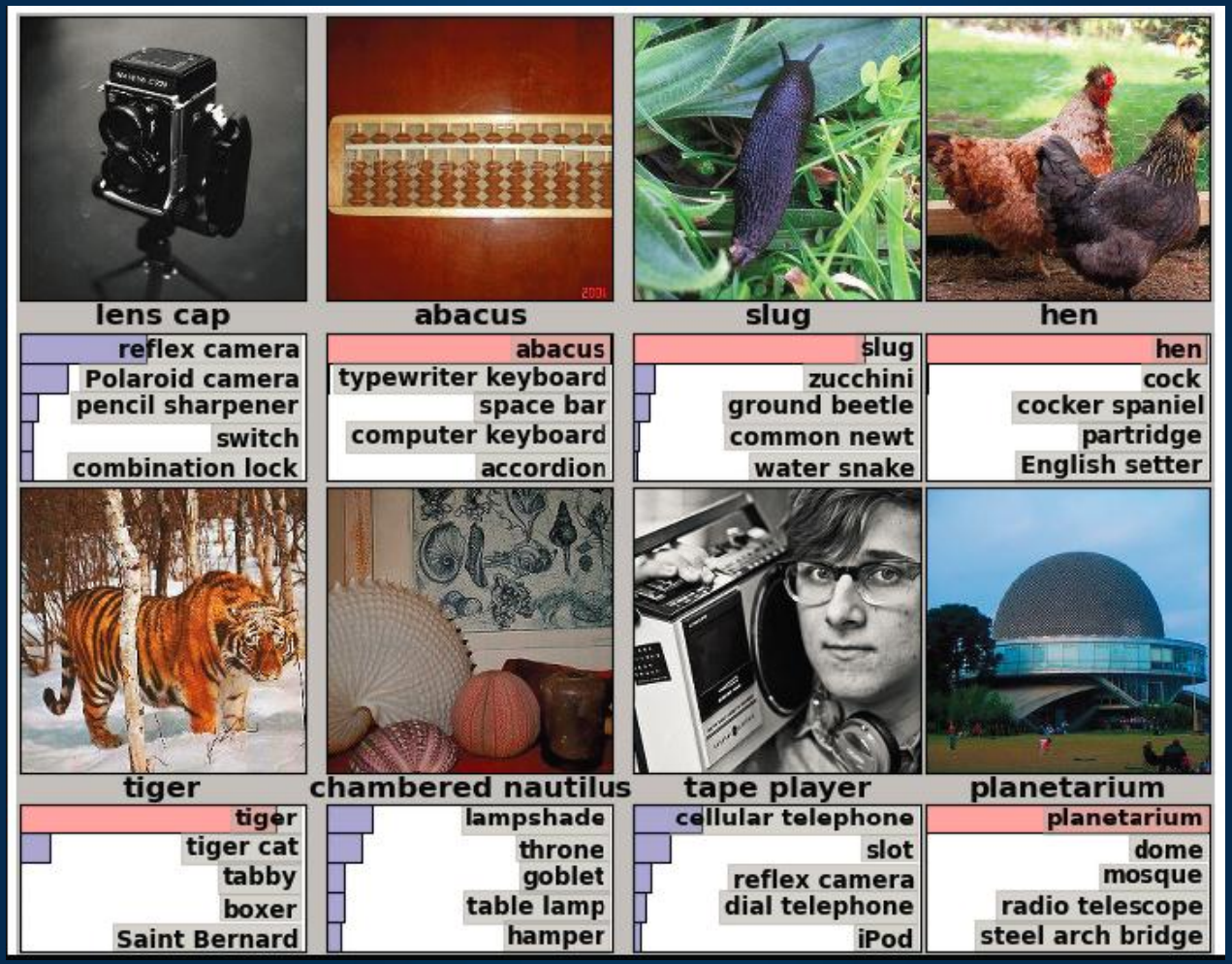
Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

CNN story: 2012 - ILSVRC

Imagenet data base: 14 mln labeled images, 20K categories

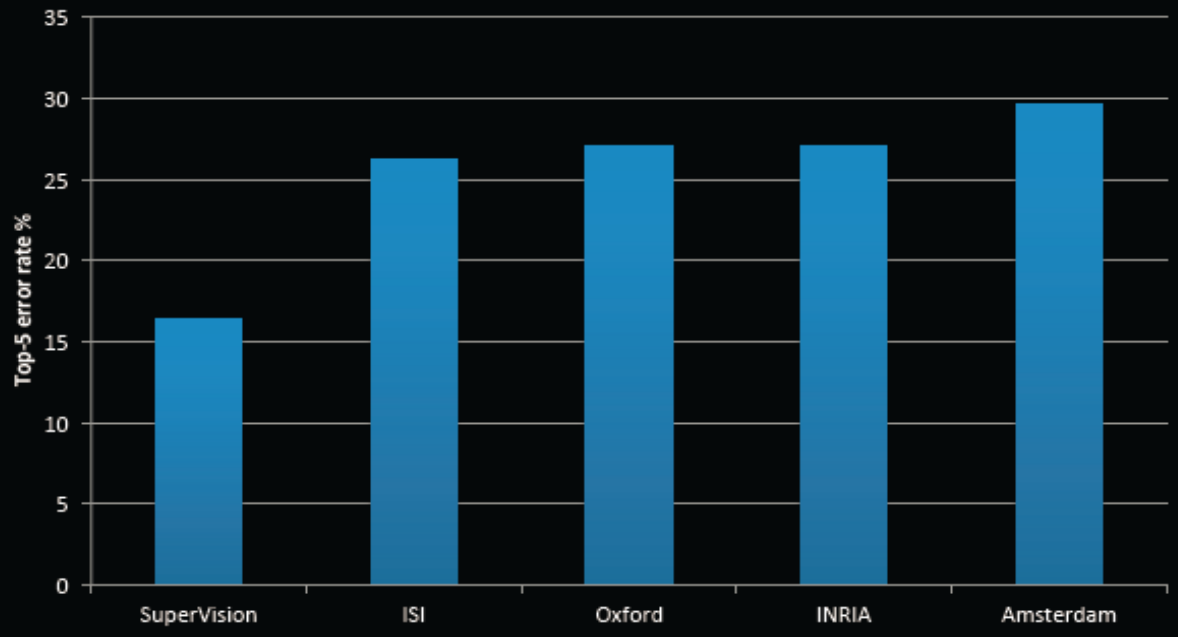


ILSVRC: Classification



Imagenet Classifications 2012

- Krizhevsky et al. -- 16.4% error (top-5)
- Next best (non-convnet) – 26.2% error



ILSVRC 2012: top rankers

http://www.image-net.org/challenges/LSVRC/2012/results.html

N	Error-5	Algorithm	Team	Authors
1	0.153	Deep Conv. Neural Network	Univ. of Toronto	Krizhevsky et al
2	0.262	Features + Fisher Vectors + Linear classifier	ISI	Gunji et al
3	0.270	Features + FV + SVM	OXFORD_VGG	Simonyan et al
4	0.271	SIFT + FV + PQ + SVM	XRCE/INRIA	Perronin et al
5	0.300	Color desc. + SVM	Univ. of Amsterdam	van de Sande et al

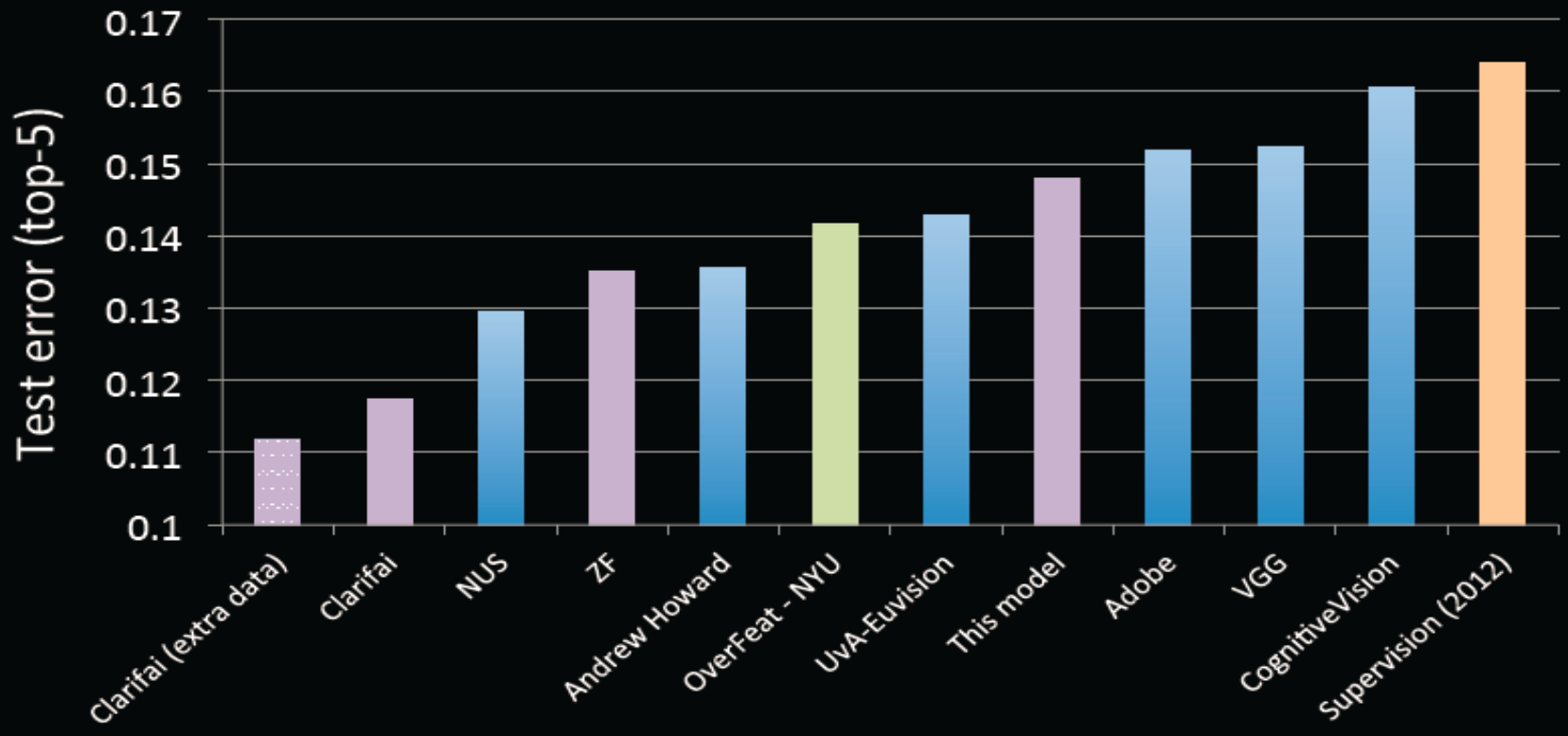
Imagenet 2013: top rankers

http://www.image-net.org/challenges/LSVRC/2013/results.php

N	Error-5	Algorithm	Team	Authors
1	0.117	Deep Convolutional Neural Network	Clarifi	Zeiler
2	0.129	Deep Convolutional Neural Networks	Nat.Univ Singapore	Min LIN
3	0.135	Deep Convolutional Neural Networks	NYU	Zeiler Fergus
4	0.135	Deep Convolutional Neural Networks		Andrew Howard
5	0.137	Deep Convolutional Neural Networks	Overfeat NYU	Pierre Sermanet et al

Imagenet Classifications 2013

- http://www.image-net.org/challenges/LSVRC/2013/results.php



- Pre-2012: 26.2% error → 2012: 16.5% error → 2013: 11.2% error

Conv Net Topology

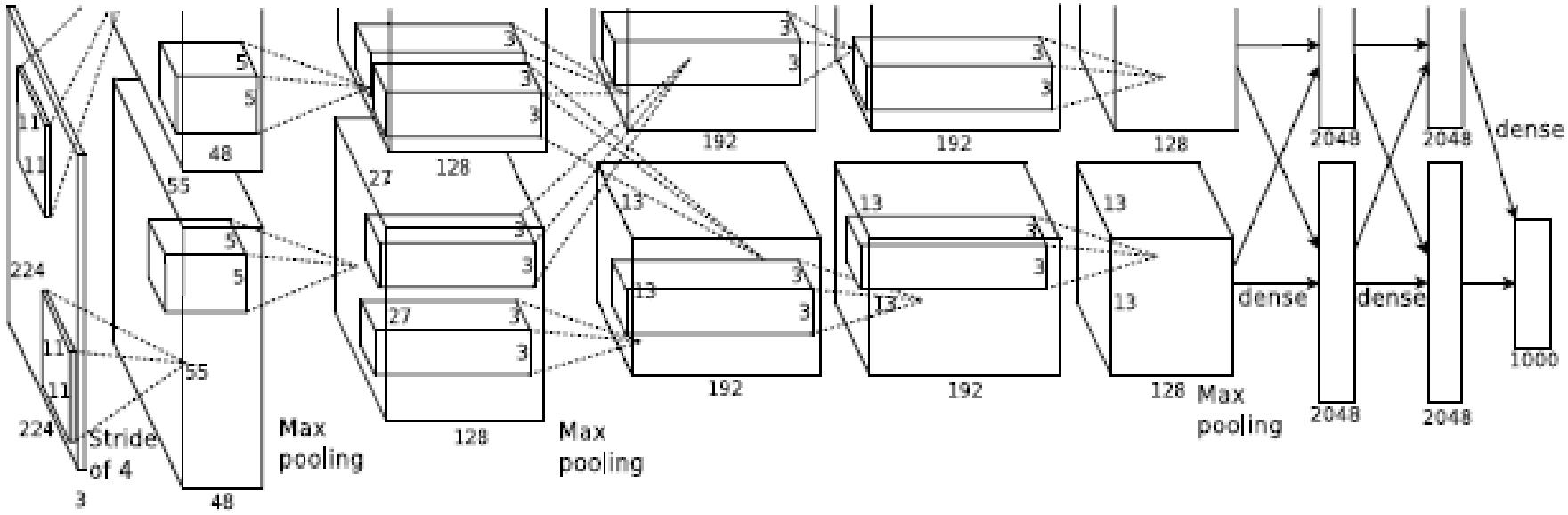
- 5 convolutional layers
- 3 fully connected layers + soft-max
- 650K neurons , 60 Mln weights

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

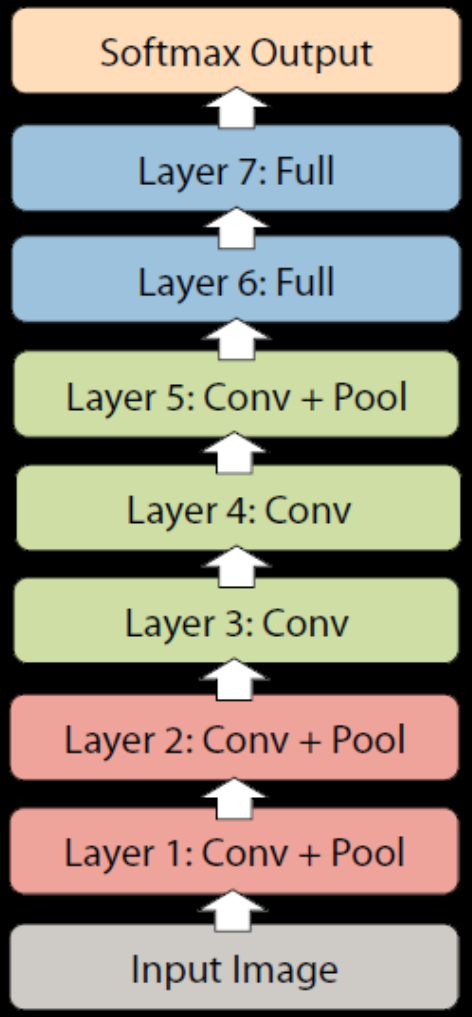
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



Why ConvNet should be Deep?

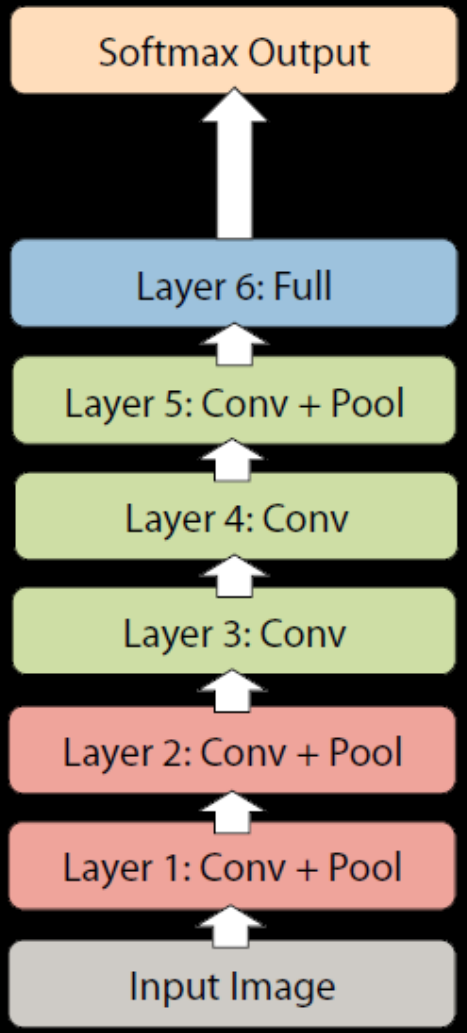
- 8 layers total
- Trained on Imagenet dataset [Deng et al. CVPR'09]
- 18.2% top-5 error
- Our reimplementation:
  18.1% top-5 error



Rob Fergus, NIPS 2013

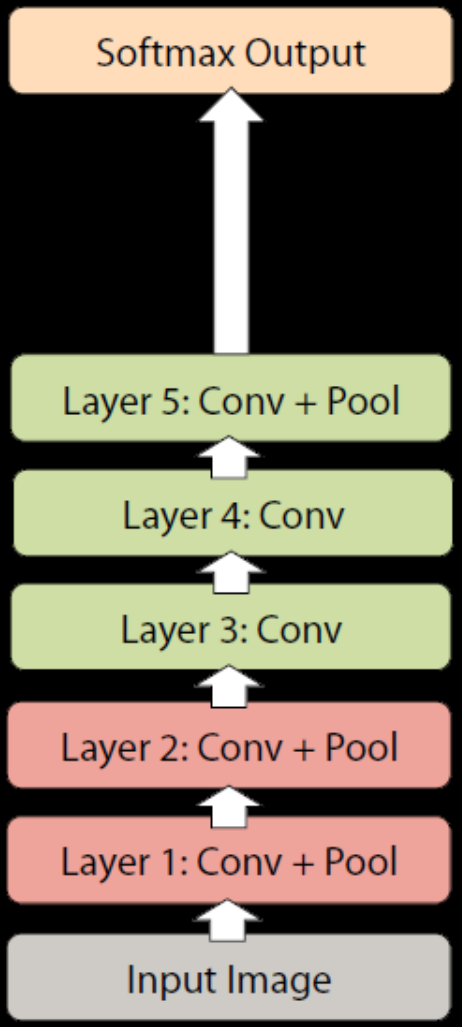
Why ConvNet should be Deep?

- Remove top fully connected layer
  - Layer 7
- Drop 16 million parameters
- Only 1.1% drop in performance!



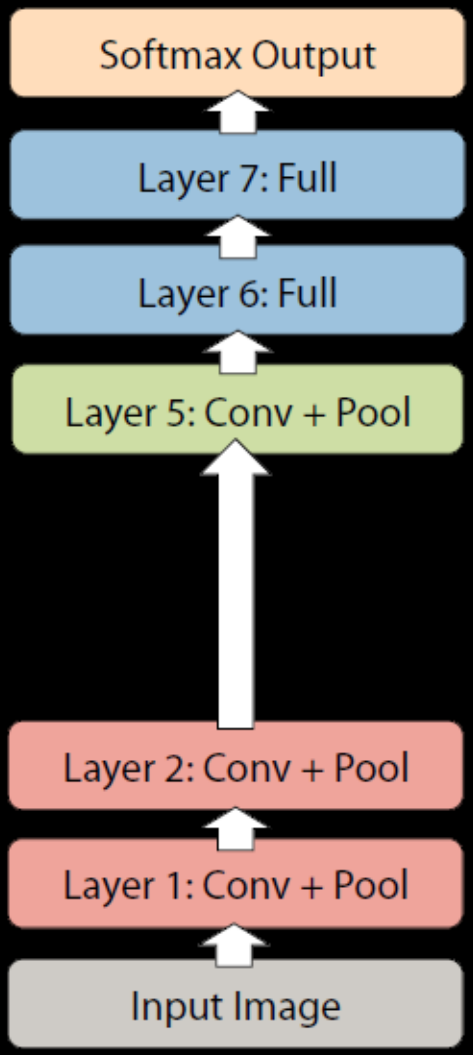
Why ConvNet should be Deep?

- Remove both fully connected layers
  - Layer 6 & 7
- Drop ~50 million parameters
- 5.7% drop in performance



Why ConvNet should be Deep?

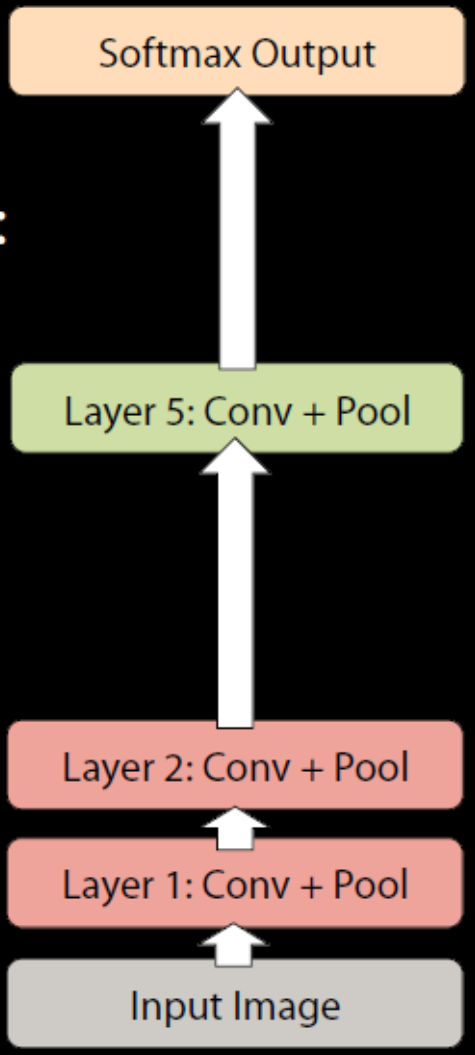
- Now try removing upper feature extractor layers:
  - Layers 3 & 4
- Drop ~1 million parameters
- 3.0% drop in performance



Why ConvNet should be Deep?

- Now try removing upper feature extractor layers & fully connected:
  - Layers 3, 4, 6 ,7
- Now only 4 layers
- 33.5% drop in performance

→Depth of network is key



Conv Nets:
beyond Visual Classification

CNN applications

CNN is a big hammer

Plenty low hanging fruits

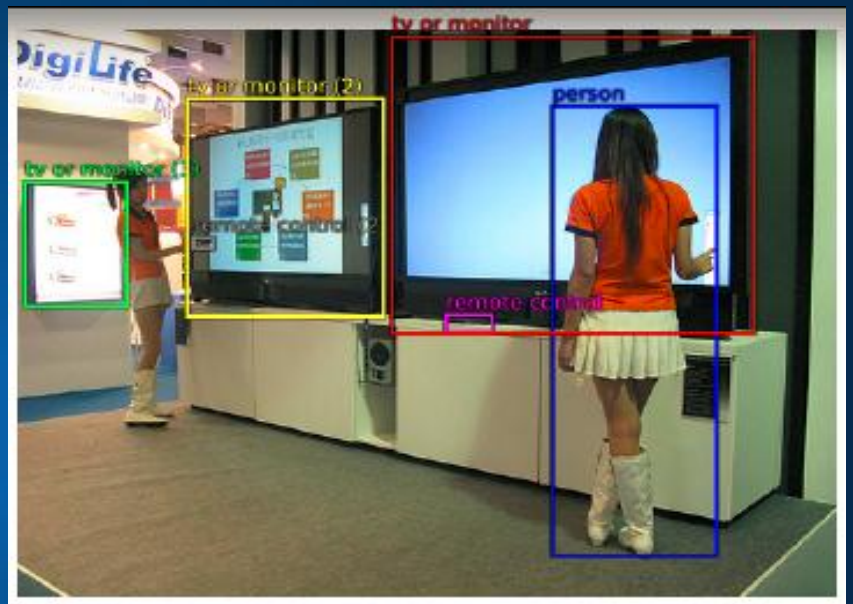
You need just a right nail!

ICRI-CI Intel Collaborative Research Institute Computational Intelligence

intel

Conv NN: Detection

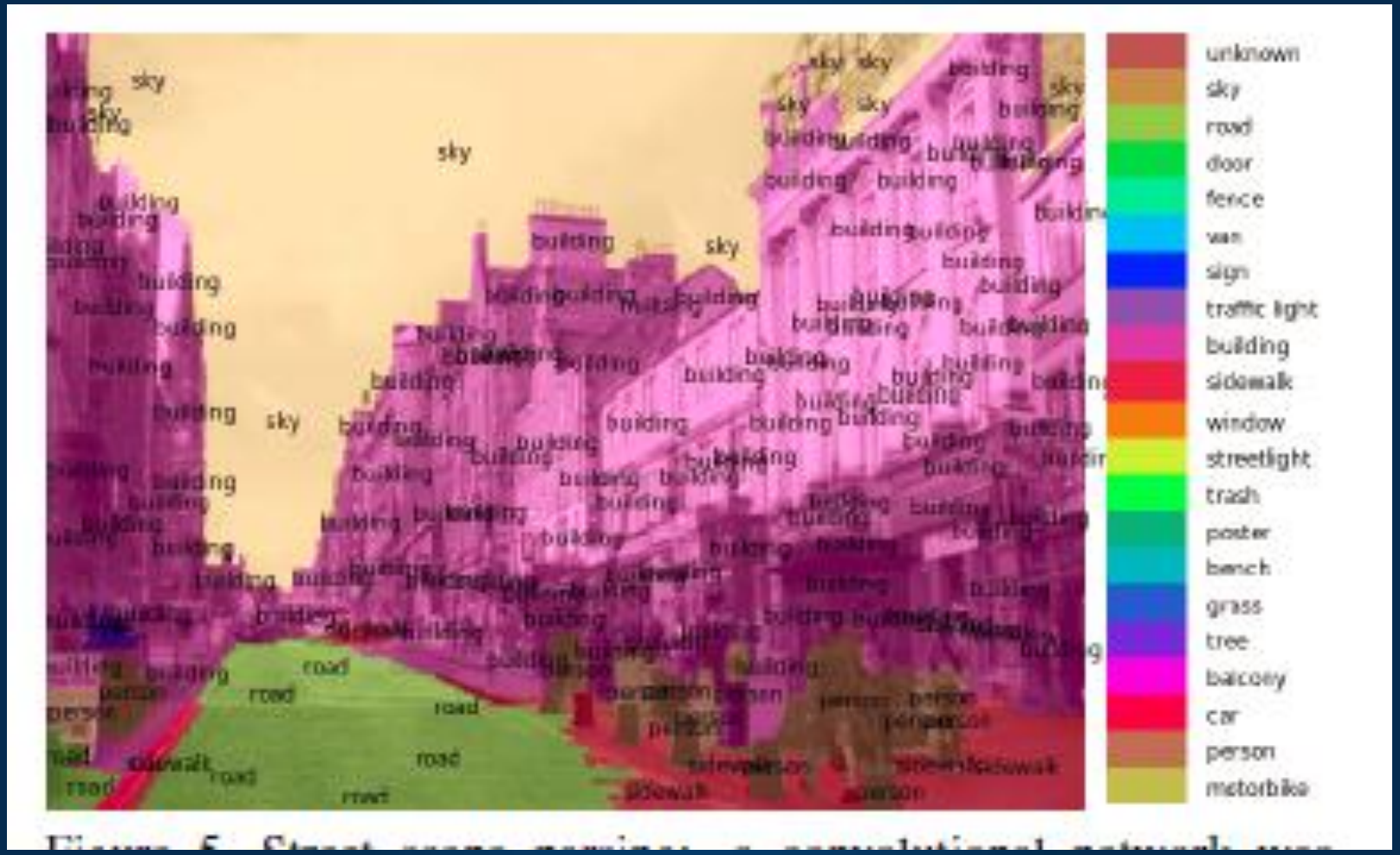
Groundtruth:
strawberry
strawberry (2)
strawberry (3)
strawberry (4)
strawberry (5)
strawberry (6)
strawberry (7)
strawberry (8)
strawberry (9)
strawberry (10)
apple
apple (2)
apple (3)



Groundtruth:
tv or monitor
tv or monitor (2)
tv or monitor (3)
person
remote control
remote control (2)

Sermanet, CVPR 2014

Conv NN: Scene parsing



Farabet, PAMI 2013

CNN: indoor semantic labeling RGBD

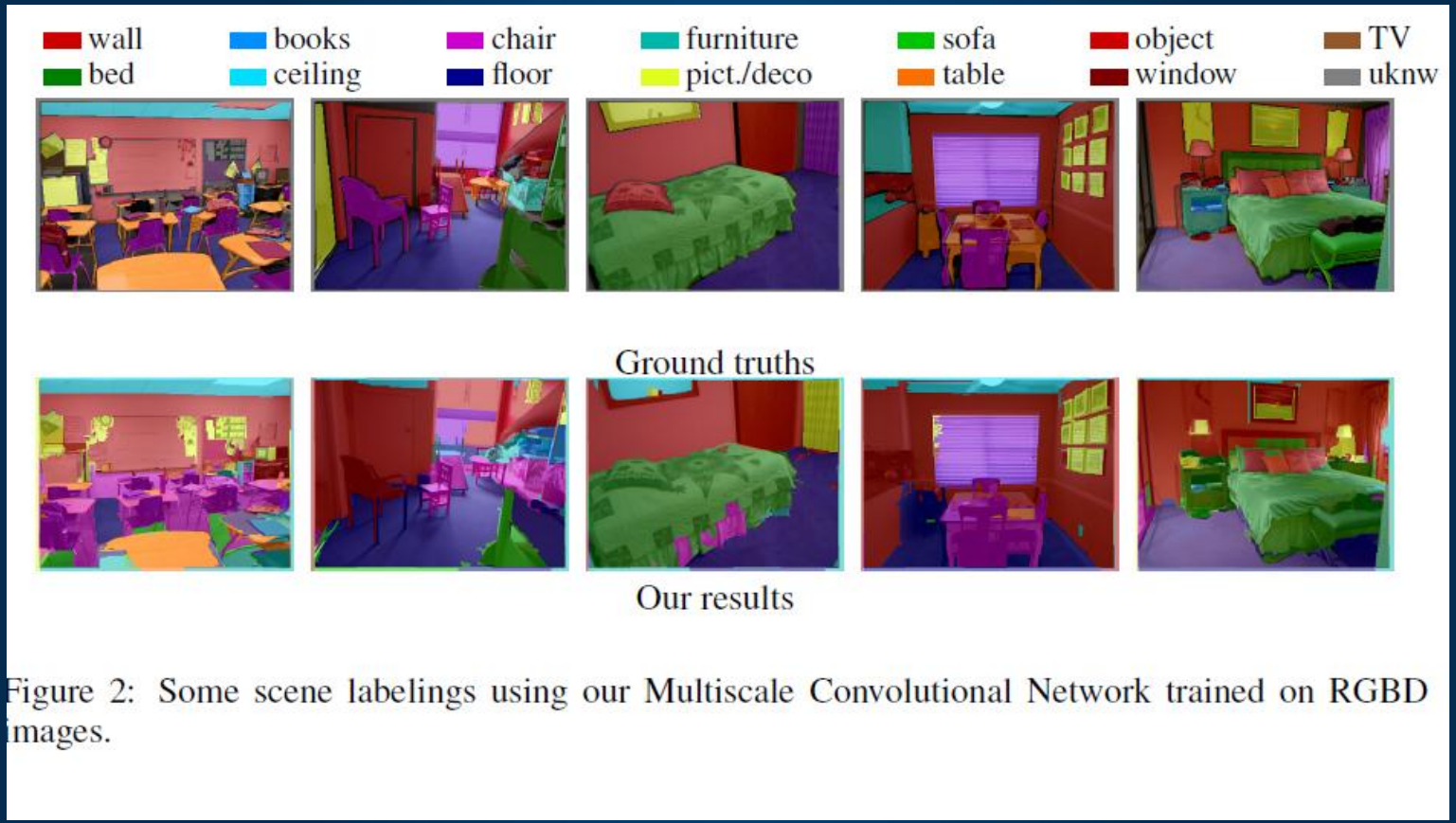


Figure 2: Some scene labelings using our Multiscale Convolutional Network trained on RGBD images.

Farabet, 2013

Conv NN: Action Detection

Taylor, ECCV 2010

Conv NN: Image Processing

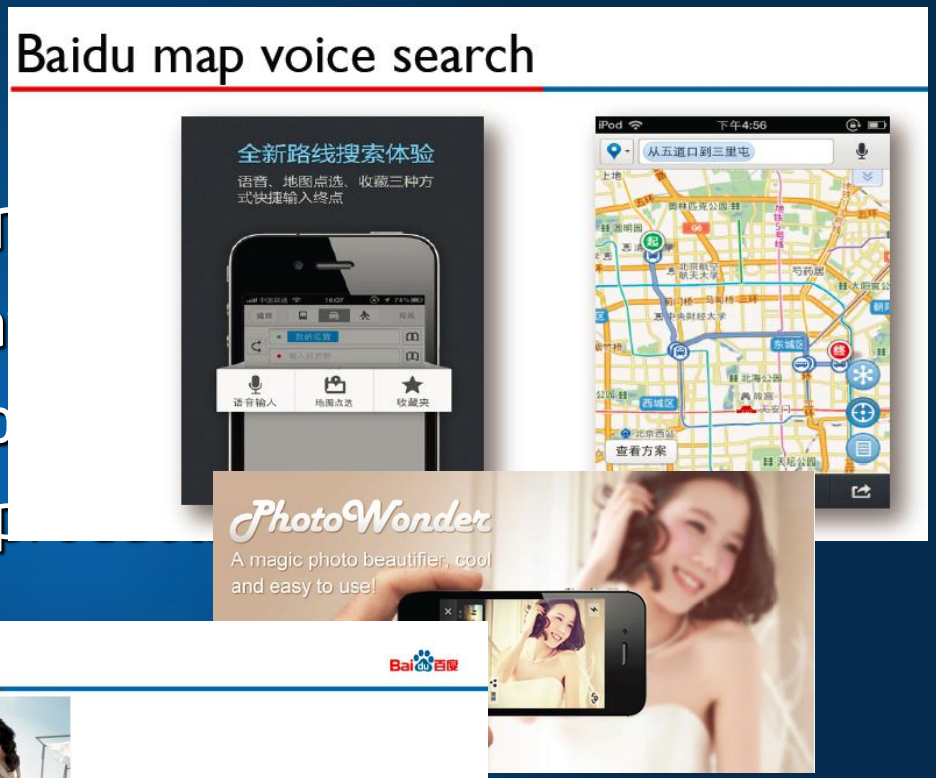
Eigen , ICCV 2010

BACKUP

BUZZ

A lot of buzz about Deep Learning

- July 2012 - Started DL lab
- Nov 2012- Big improver
  - Speech – reduce Error Ra
  - OCR – reduce Error rate b
- 2013 launched 5 DL based p
  - Voice search
  - Photo W
  - Visual se



A lot of buzz about Deep Learning

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

Microsoft On Deep Learning for Speech goto 3:00-5:10

A lot of buzz about Deep Learning

Why Google invest in Deep Learning

A lot of buzz about Deep Learning

Courant's LeCun to Lead Facebook's New Artificial Intelligence Group

December 9, 2013
143

Facebook has named New York University Professor Yann LeCun the director of a new laboratory devoted to research in artificial intelligence and deep learning.

"As one of the most respected thinkers in this field, Yann has done groundbreaking rese deep learning and computer vision," said Mike Schroepfer, Facebook's chief technology "We're thrilled to welcome him to Facebook."

Facebook is building the team across three locations: Facebook's headquarters in Menlo Calif., New York City, and London.

Machine learning is a branch of artificial intelligence that involves computers "learning" t extract knowledge from massive data sets and rendering informed analyses and judgme often predicting outcomes.

LeCun, a professor at NYU's Courant Institute of Mathematical Sciences, is a pioneer in t growing field. In the 1980s, LeCun proposed one of the early versions of the back-propa algorithm, the most popular method for training artificial neural networks. In the late 19 and early 1990s at AT&T Bell Laboratories, he developed the convolutional network mod

ENTERPRISE | research | uncategorized

Facebook Taps 'Deep Learning' Giant for New AI Lab

BY CADE METZ 12.09.13 3:14 PM
Follow @cademetz

Share 239 Tweet 214 +1 43 Share 48 Pin it

Yann LeCun. *Photo: Josh Valcarcel/WIRED*

Facebook is building a research lab dedicated to the new breed of artificial intelligence, after hiring one of the preeminent researchers in the field: New York University professor Yann LeCun

NYU "Deep Learning" Professor LeCun Will Head Facebook's New Artificial Intelligence Lab, Dec 10, 2013