These CSCI-561 slides adapted from:

Introduction:
Convolutional Neural Networks for Visual Recognition

boris .ginzburg@intel.com

Adapted from:

CS294-129: Designing, Visualizing and Understanding Deep Neural Networks

John Canny

Fall 2016

Lecture 6: Projects and Training Neural Networks I

Based on notes from Andrej Karpathy, Fei-Fei Li, Justin Johnson

Acknowledgments

This presentation is heavily based on:

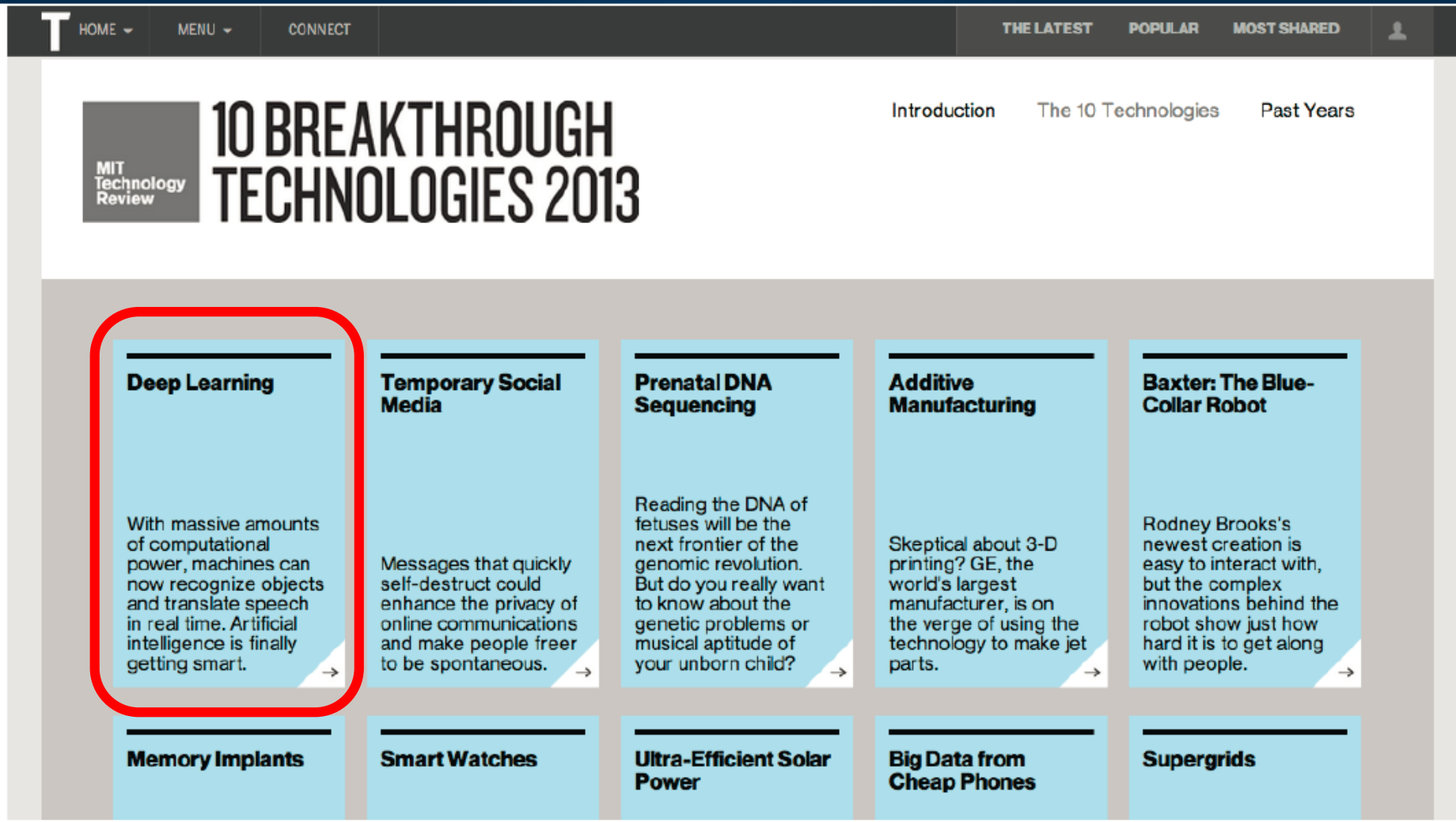
- http://cs.nyu.edu/~fergus/pmwiki/pmwiki.php
- http://deeplearning.net/reading-list/tutorials/
- http://deeplearning.net/tutorial/lenet.html
- http://ufldl.stanford.edu/wiki/index.php/UFLDL_Tutorial

… and many other

Agenda

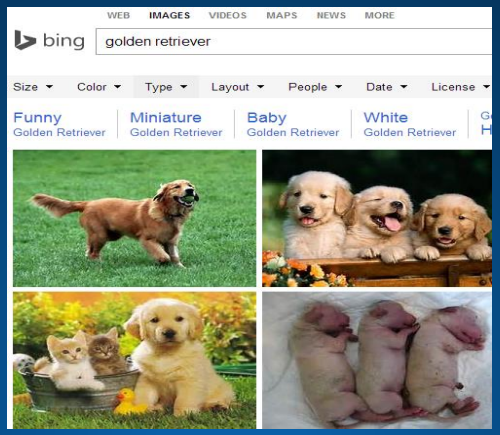
1. Overview
2. Introduction to Deep Learning
   - Classical Computer Vision vs. Deep learning
3. Introduction to Convolutional Networks
   - Basic CNN Architecture
   - Large Scale Image Classifications
   - Detection and Other Visual Apps
   - Beyond machine vision

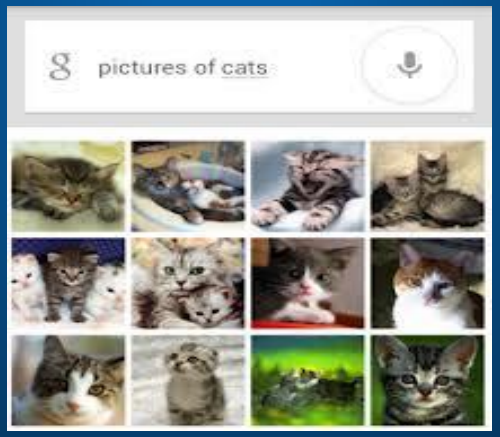
Buzz…



MIT Technology Review, April 23rd, 2013

Deep Learning – from Research to Technology





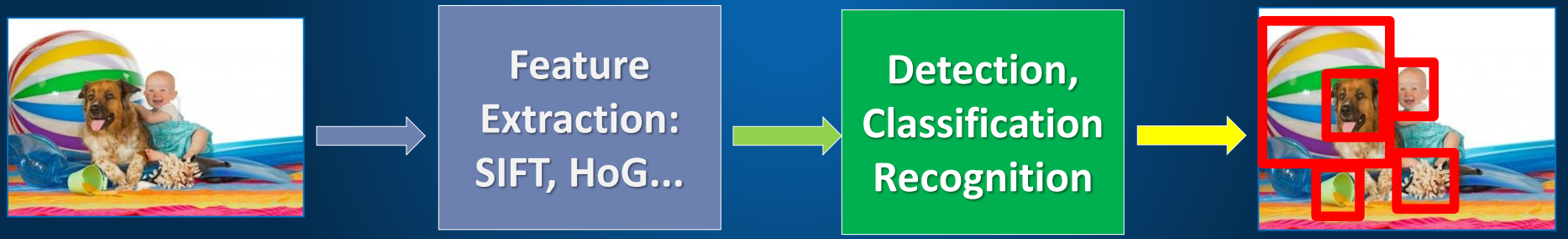
Deep Learning - breakthrough in visual and speech recognition

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



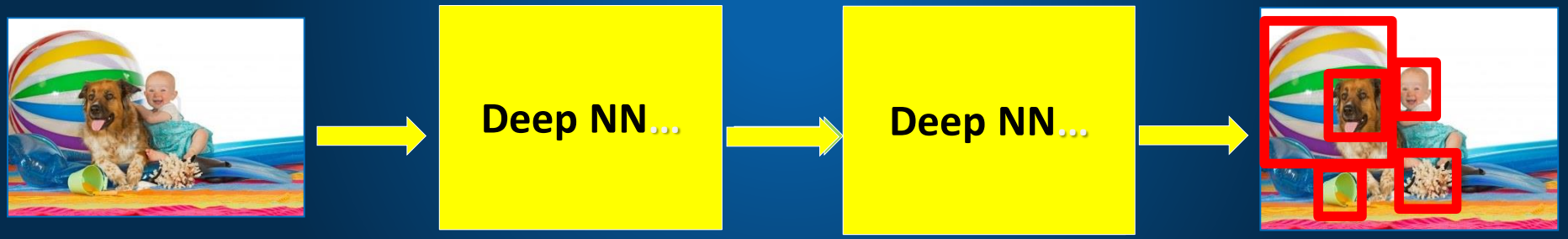
Classical CV feature definition is domain-specific and time-consuming

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

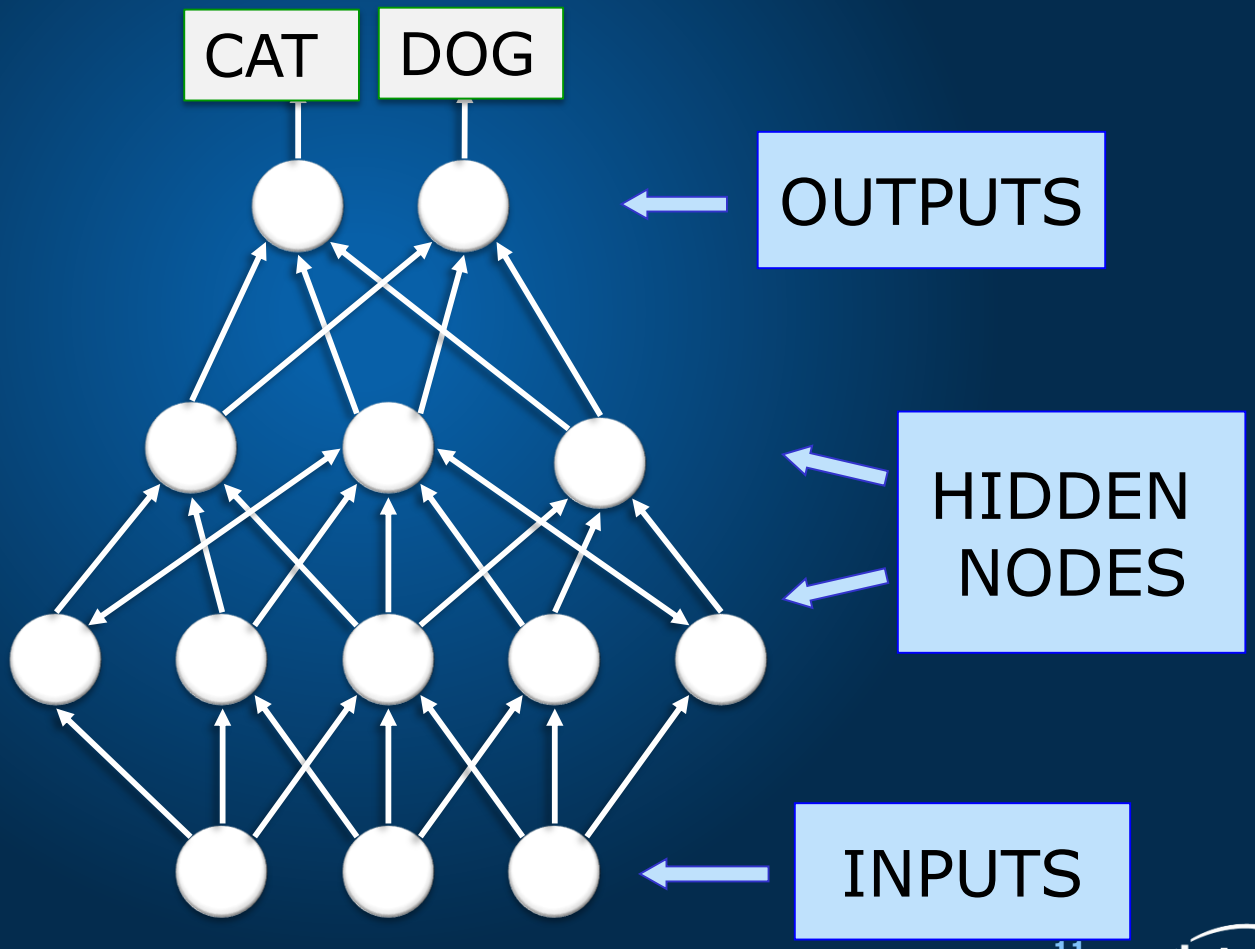
- Combine pre-defined features with learned features;
- Use best ML methods for multi-class recognition

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

Combine best of Computer Vision
Deep Learning and Machine Learning

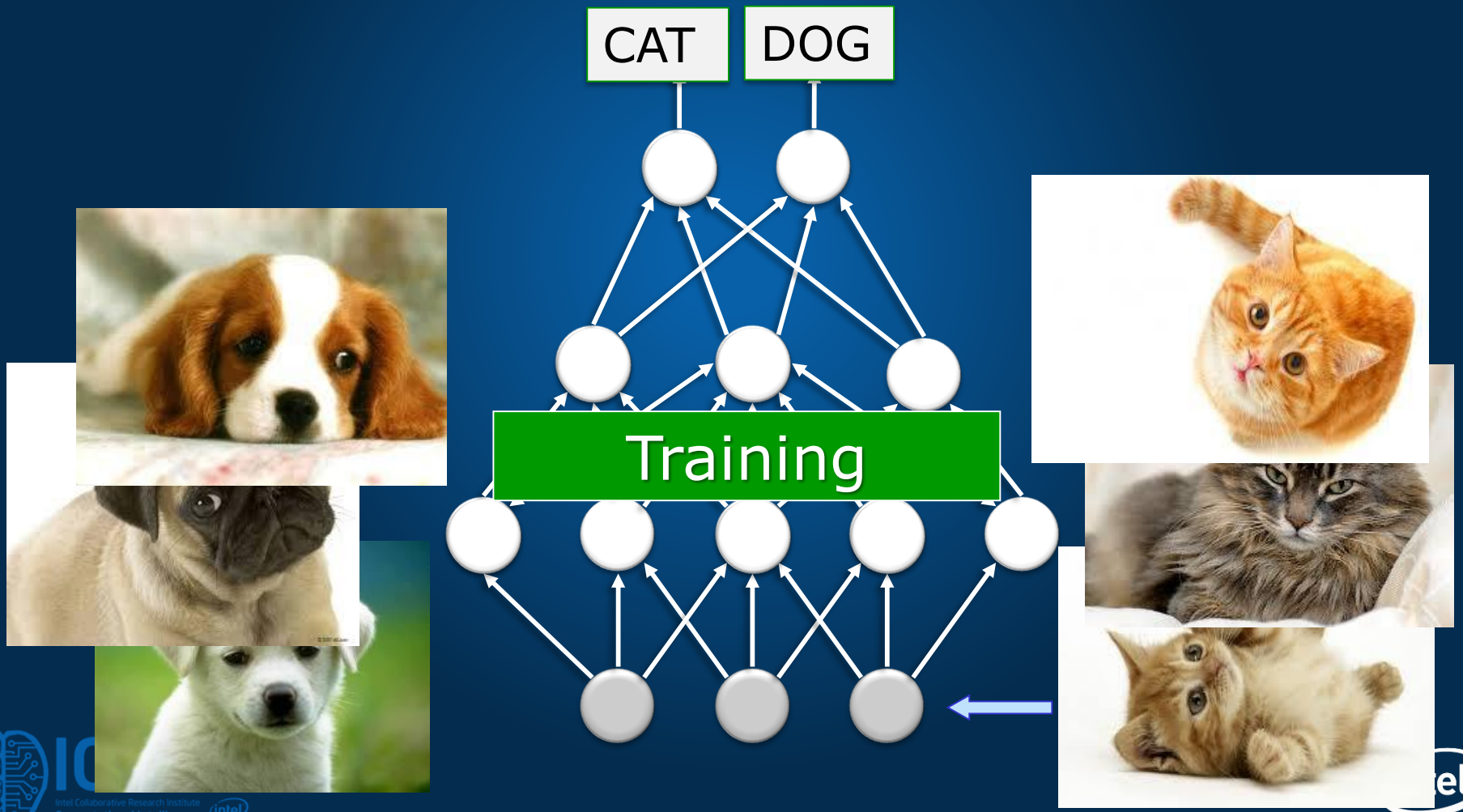
Deep Learning Basics

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



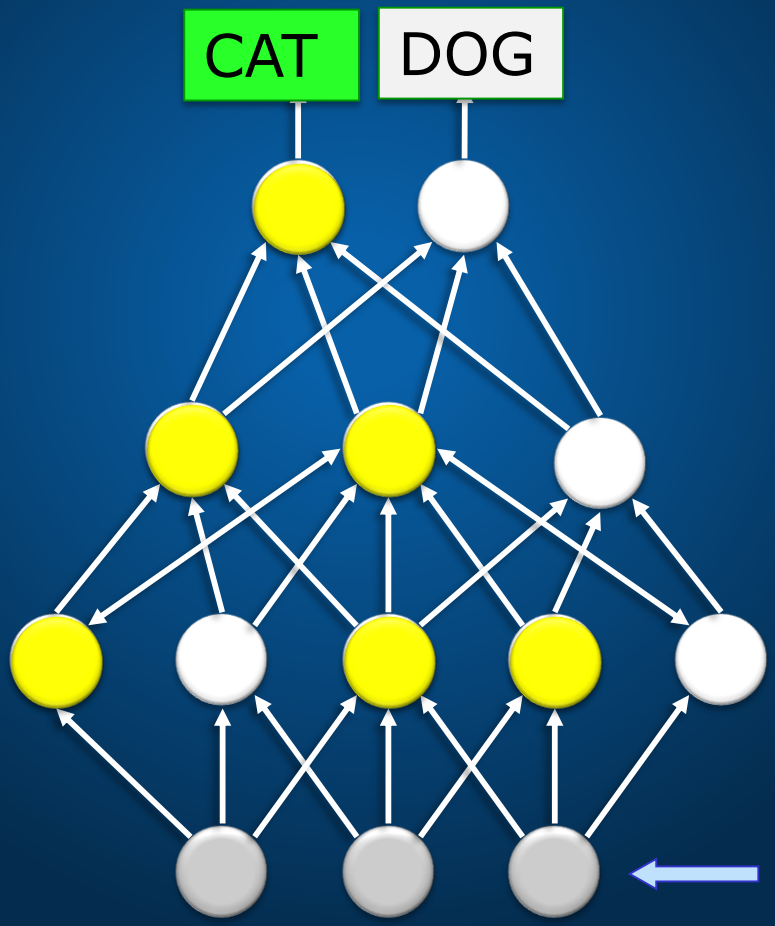
Deep Learning Basics

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



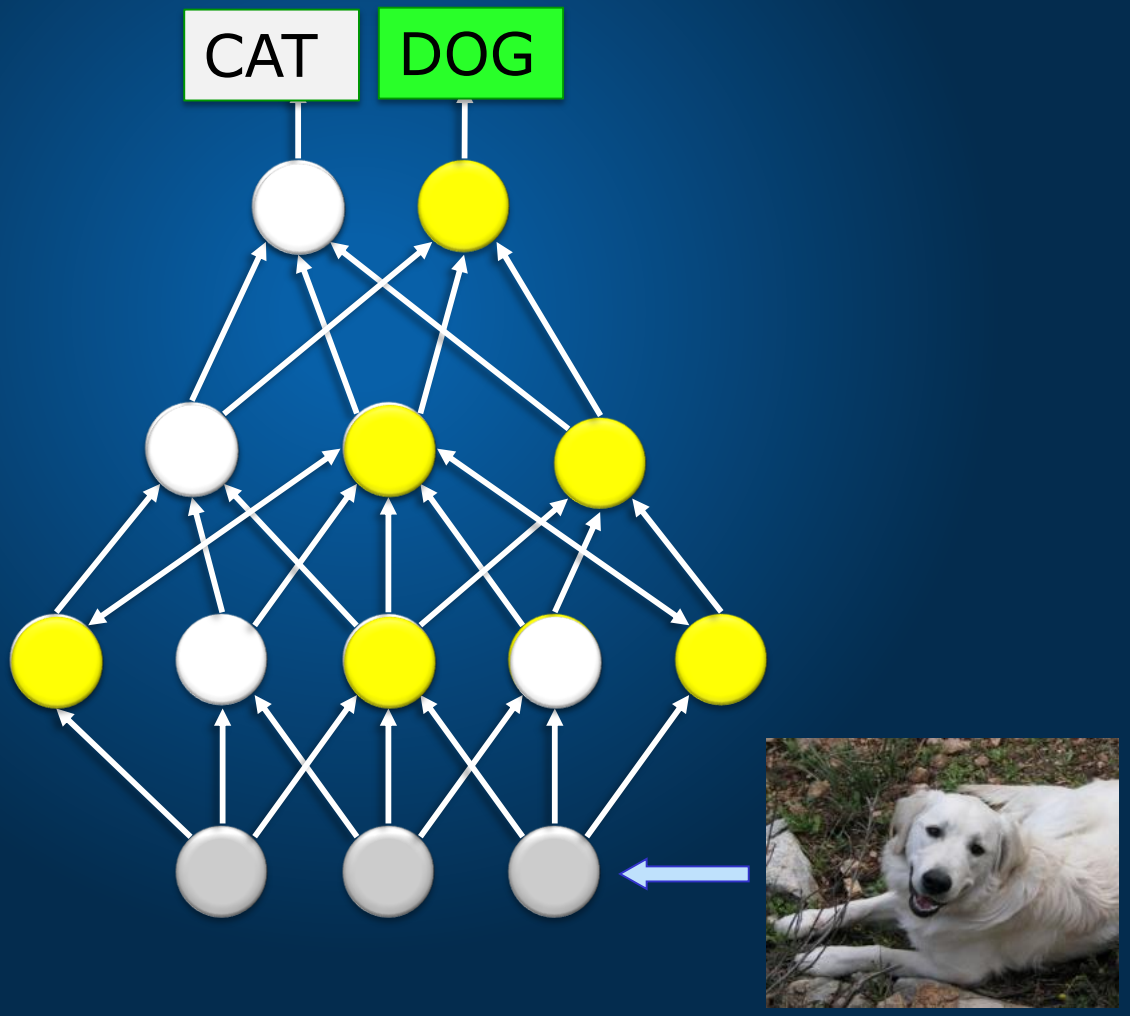
Deep Learning Basics

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



ICRI-CI
Intel Collaborative Research Institute
Computational Intelligence

intel

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:

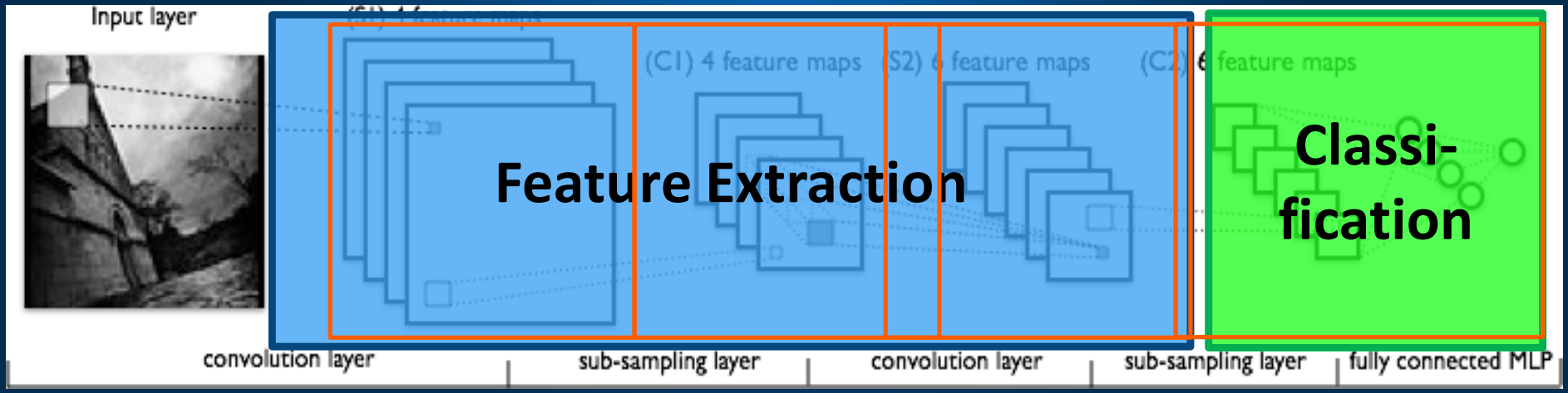
http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf

What is Convolutional NN ?

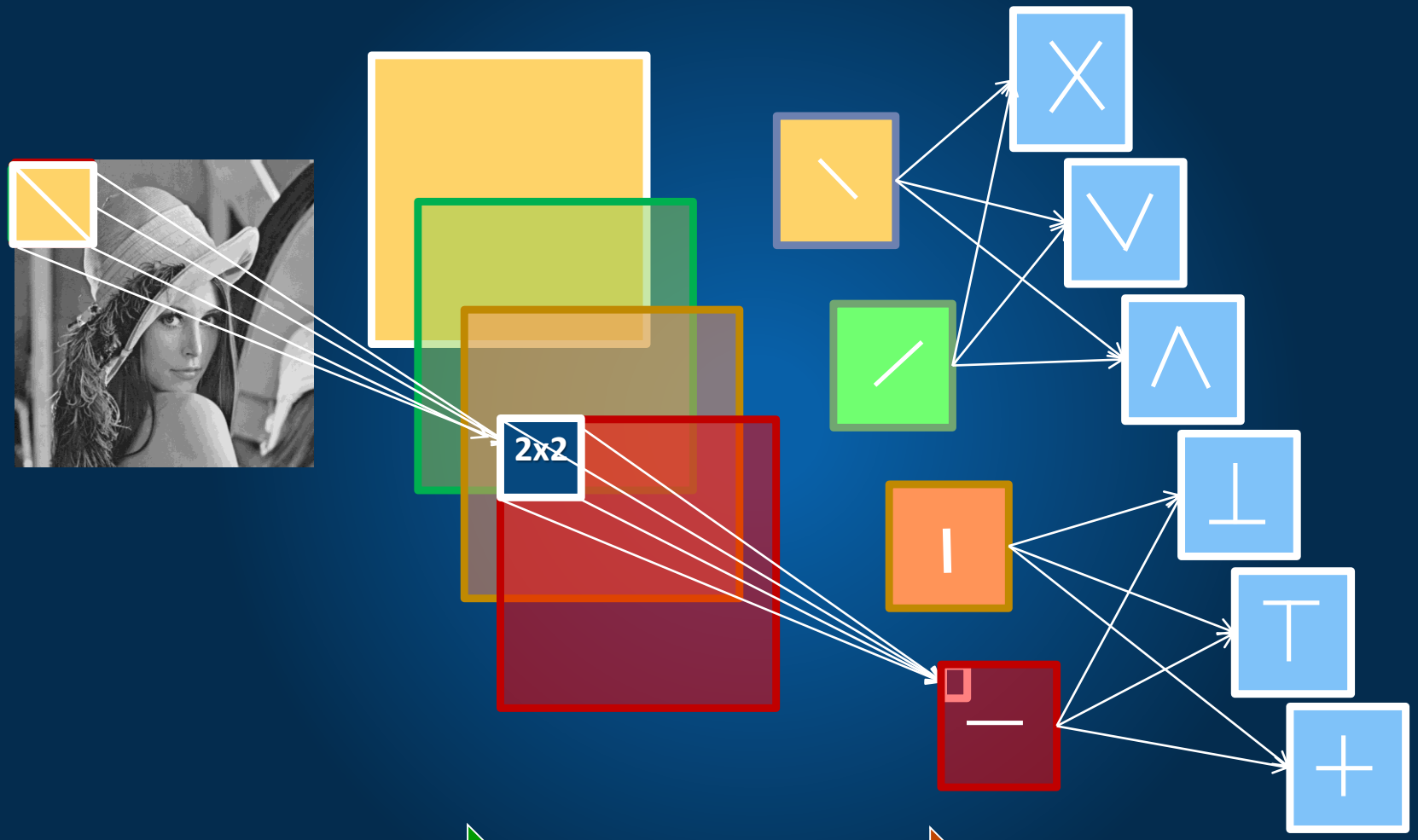
CNN - multi-layer NN architecture

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

- Supervised

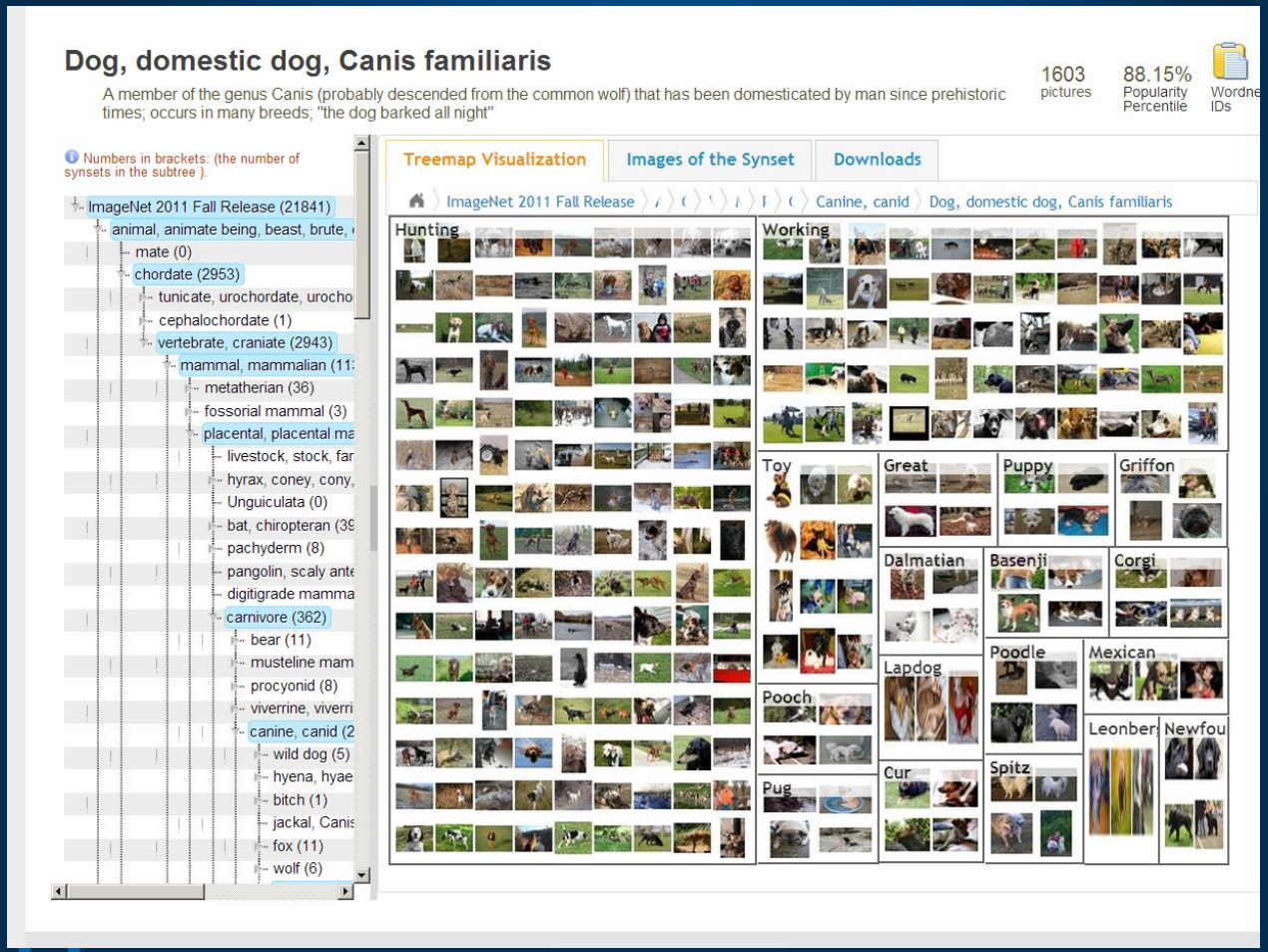


What is Convolutional NN ?

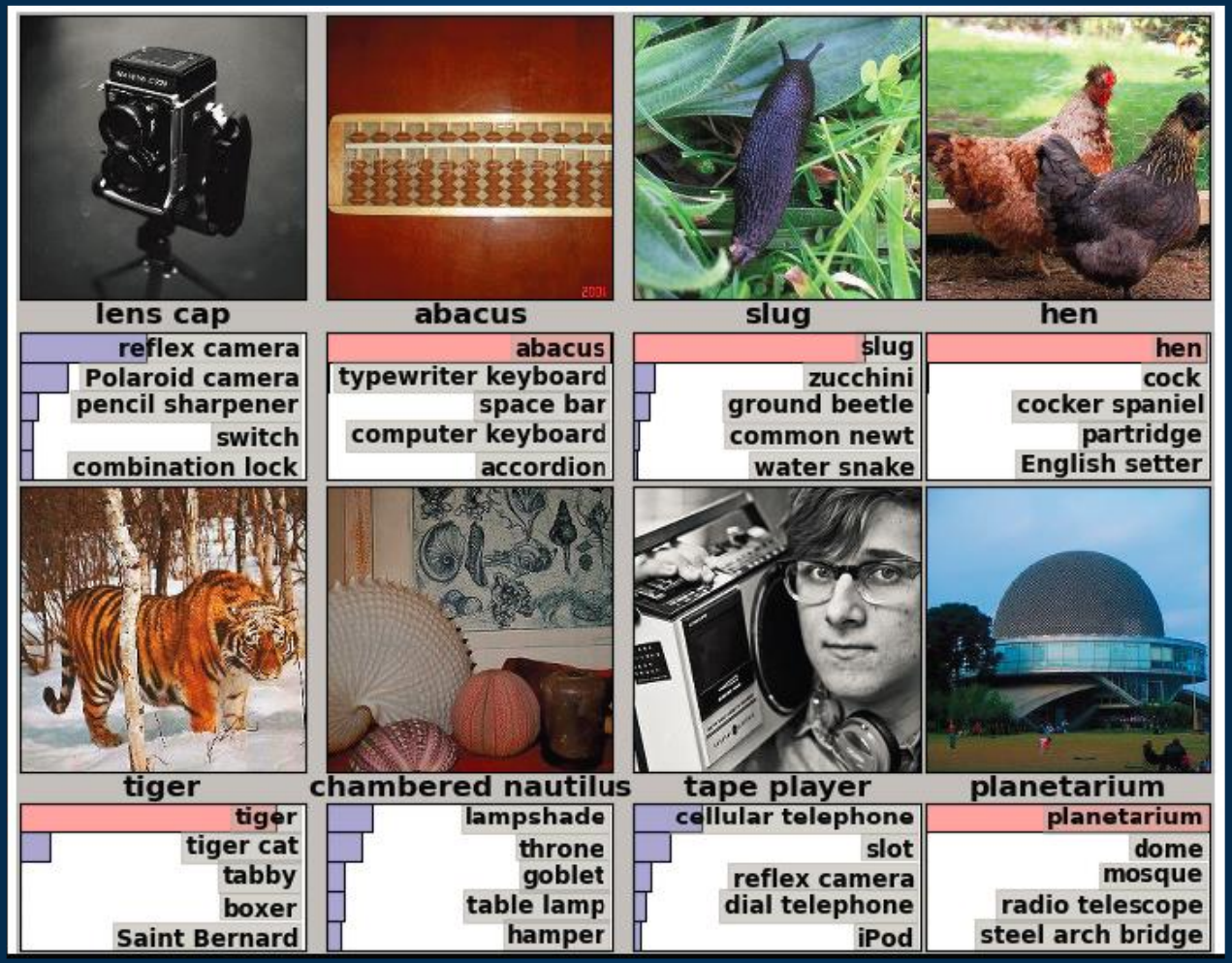


CNN success story: ILSVRC 2012

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

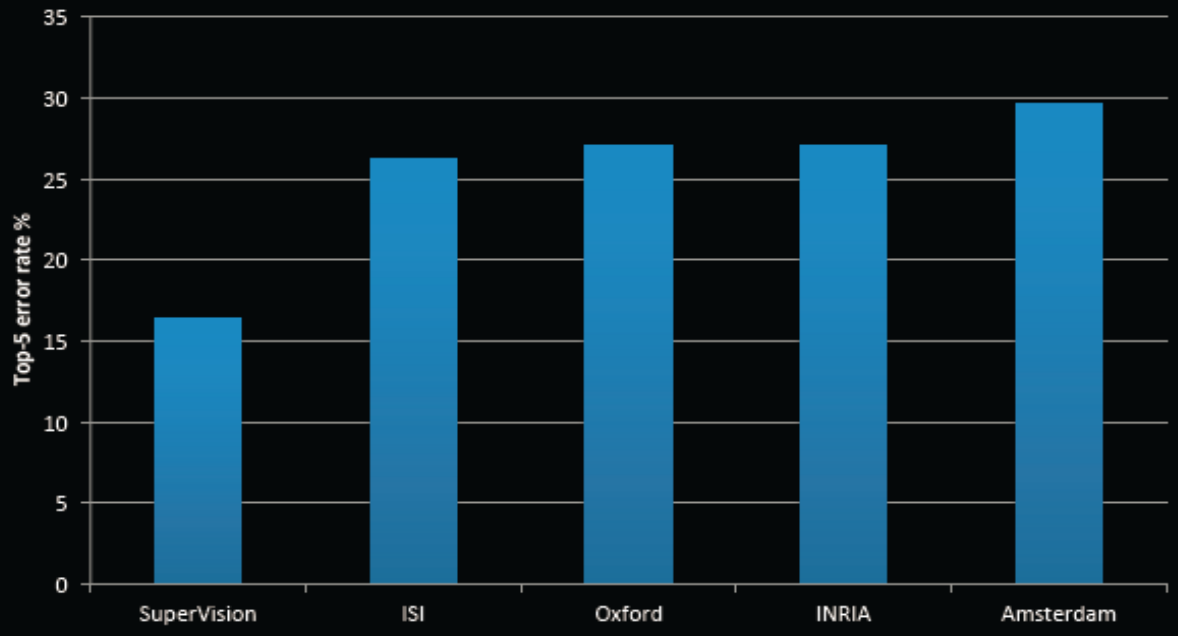


ILSVRC: Classification



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

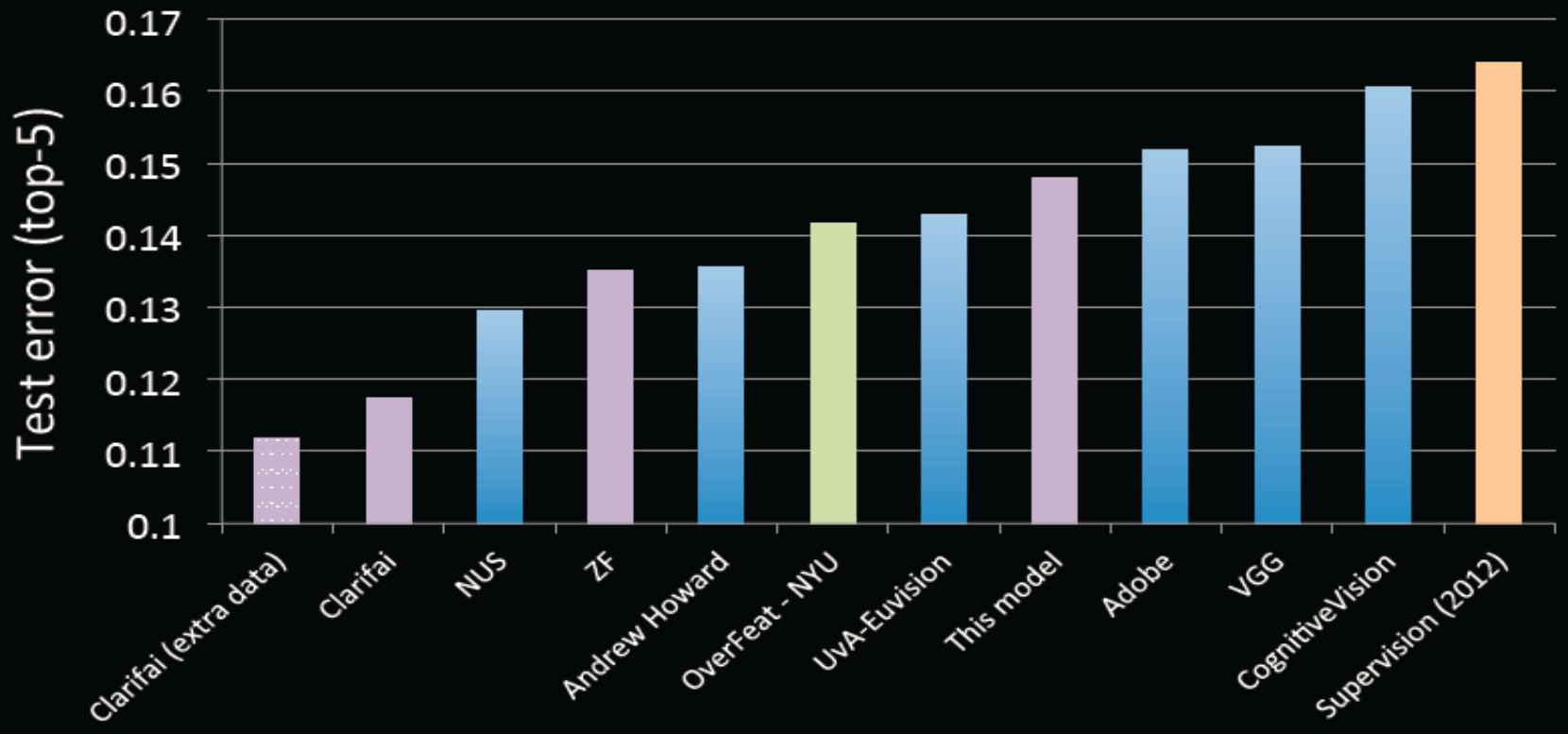
Imagenet 2013: top rankers

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

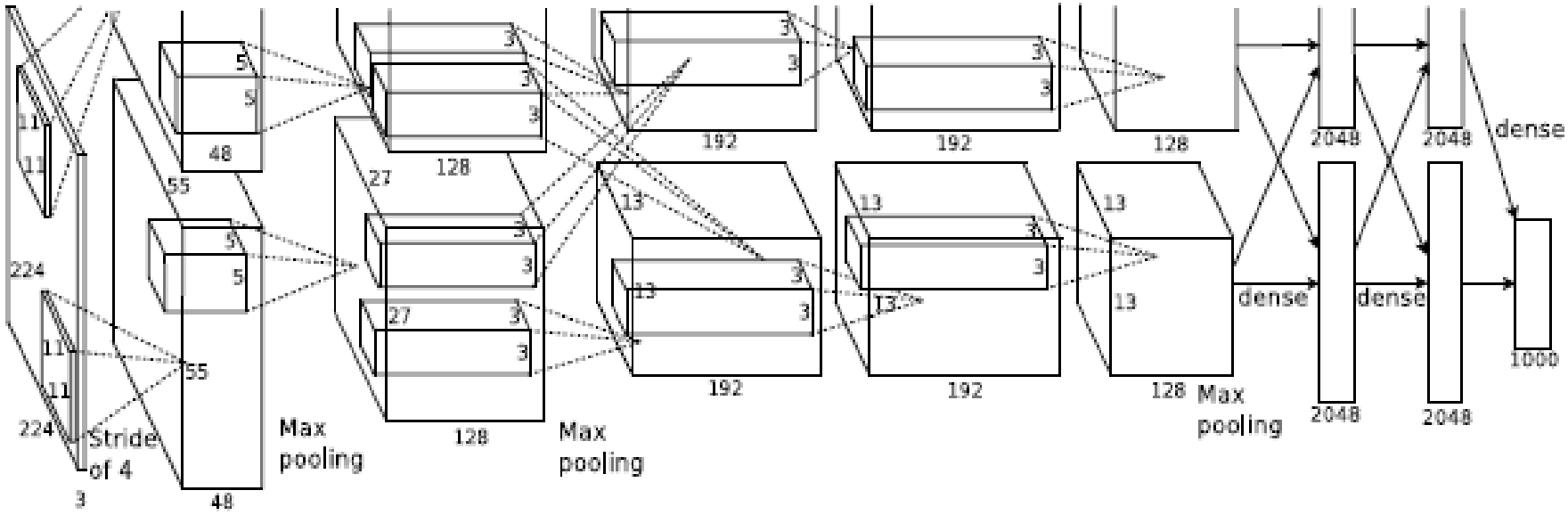
- 5 convolutional layers
- 3 fully connected layers + soft-max
- 650K neurons , 60 Mln weights

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



Conv Nets:
beyond Visual Classification

CNN applications

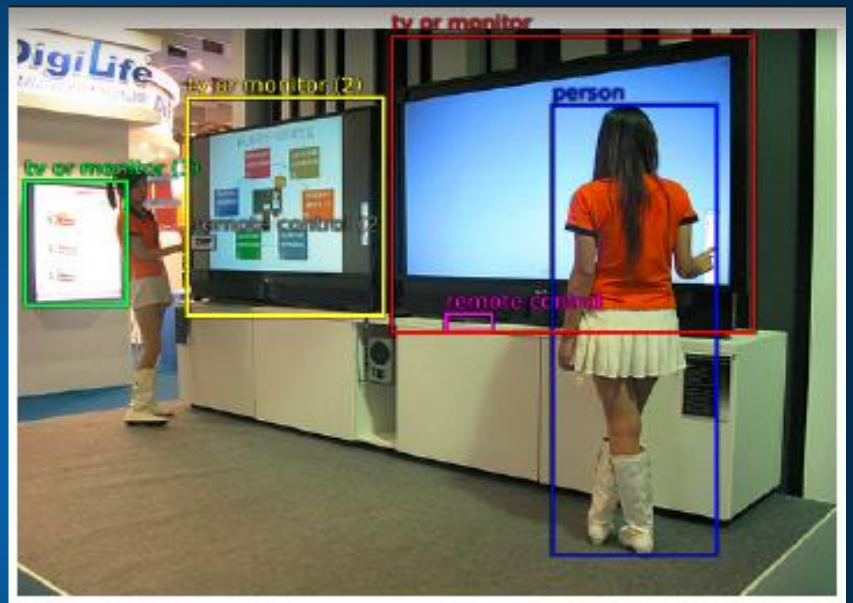
CNN is a big hammer

Plenty low hanging fruits

You need just a right nail!

Conv NN: Detection

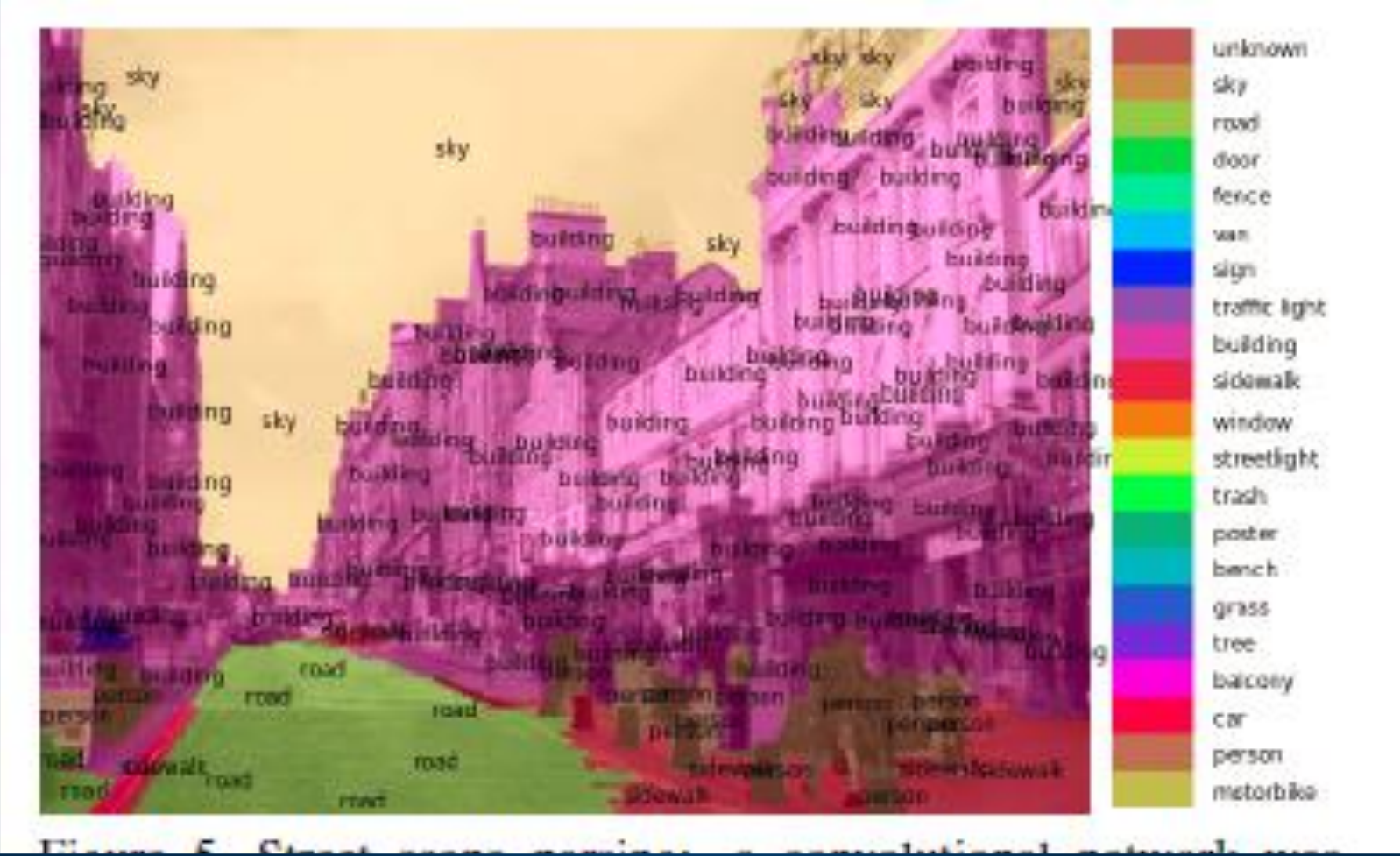
Groundtruth:
strawberry
strawberry (2)
strawberry (3)
strawberry (4)
strawberry (5)
strawberry (6)
strawberry (7)
strawberry (8)
strawberry (9)
strawberry (10)
apple
apple (2)
apple (3)



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

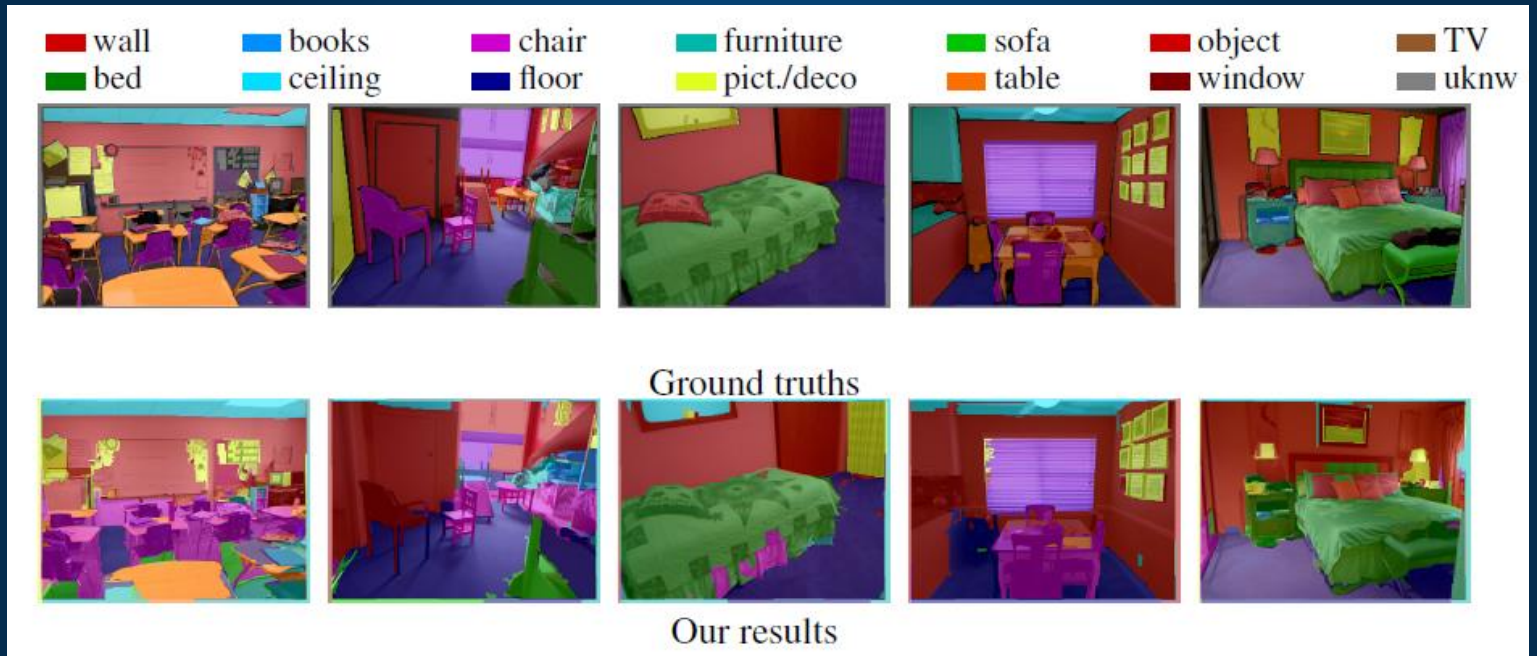


Figure 2: Some scene labelings using our Multiscale Convolutional Network trained on RGBD images.

Farabet, 2013

Conv NN: Action Detection

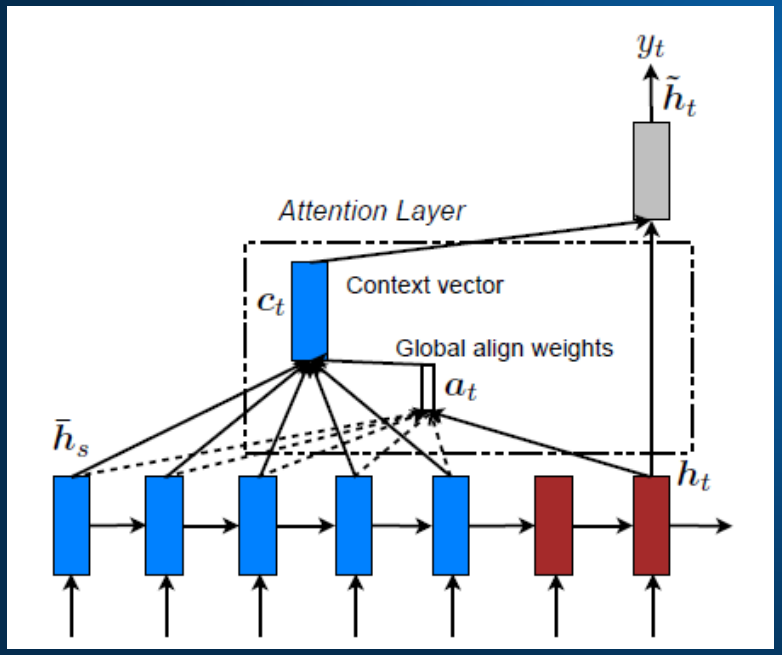
Taylor, ECCV 2010

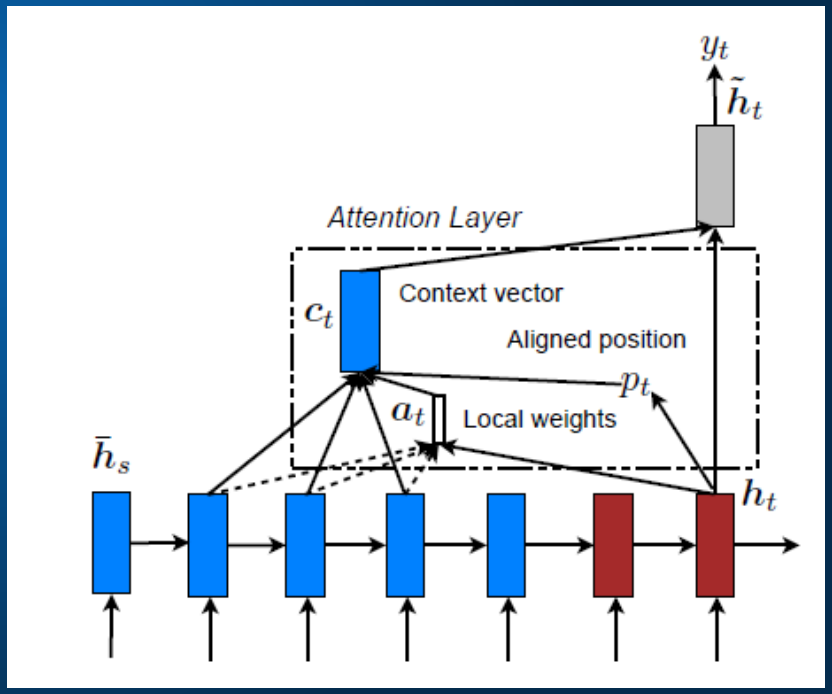
Conv NN: Image Processing

Eigen , ICCV 2010

Machine Translation

Use sequence-to-sequence models + attention.
"Effective Approaches to Attention-based Neural Machine Translation" Minh-Thang Luong, Hieu Pham, Christopher D. Manning
Datasets: see paper





End-To-End Memory Networks

End to End Memory Networks (Tensorflow versions exist here and here)
Paper: S. Sukhbaatar, A. Szlam, J. Weston, and R. Fergus. End-to-end memory networks
Dataset: https://research.facebook.com/research/babi/

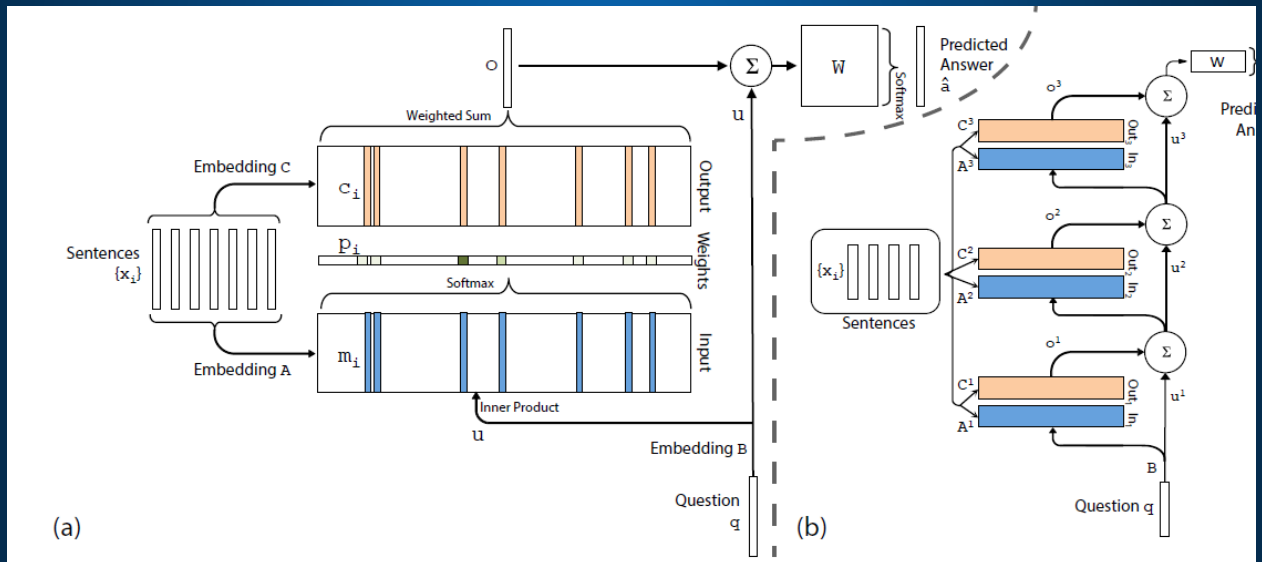
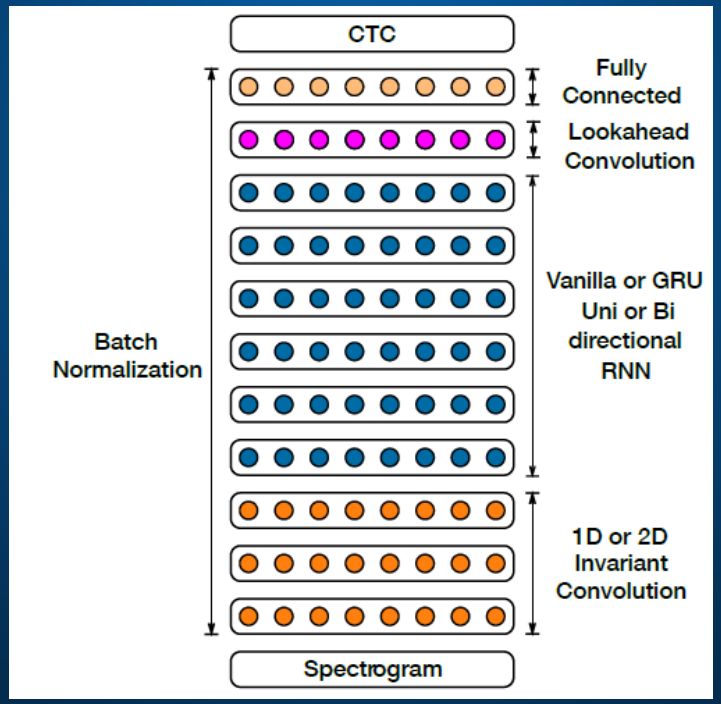


Figure 1: (a): A single layer version of our model. (b): A three layer version of our model.

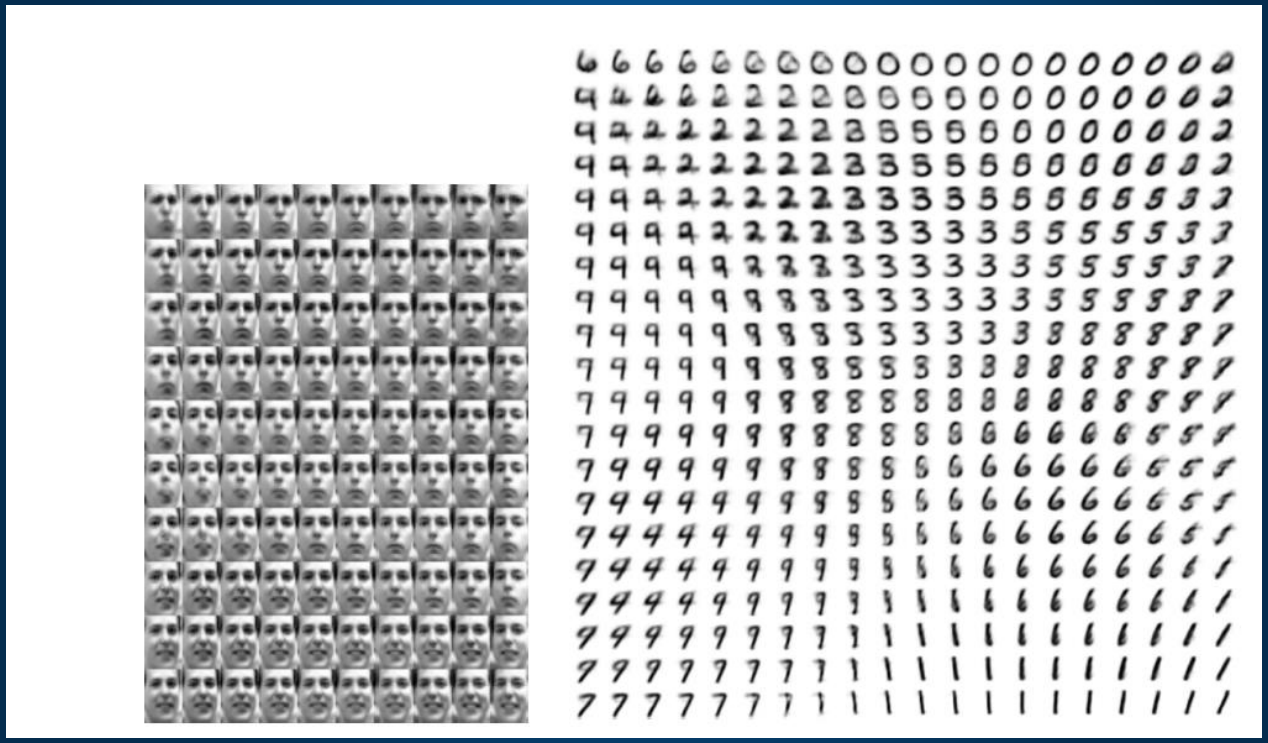
Deep Speech 2

Paper: Dario Amodei et al. "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin".
Torch code: Here and here is an intro.
Datasets: TIMIT continuous speech data, WMT `15.



Variational AutoEncoders

Paper: D. Kingma and M. Welling. "Stochastic Gradient VB and the Variational Auto-Encoder"
Dataset: MNIST, Could try CIFAR 10 or 100 or ImageNet

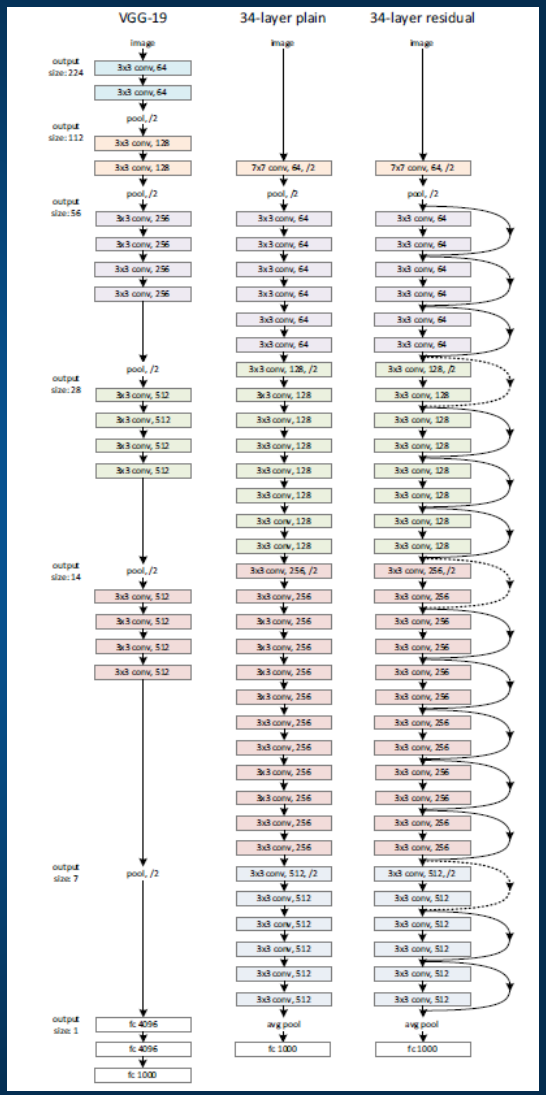


Residual Networks

Paper: Kaiming He et. al. "Deep Residual Learning for Image Recognition"

Torch code here

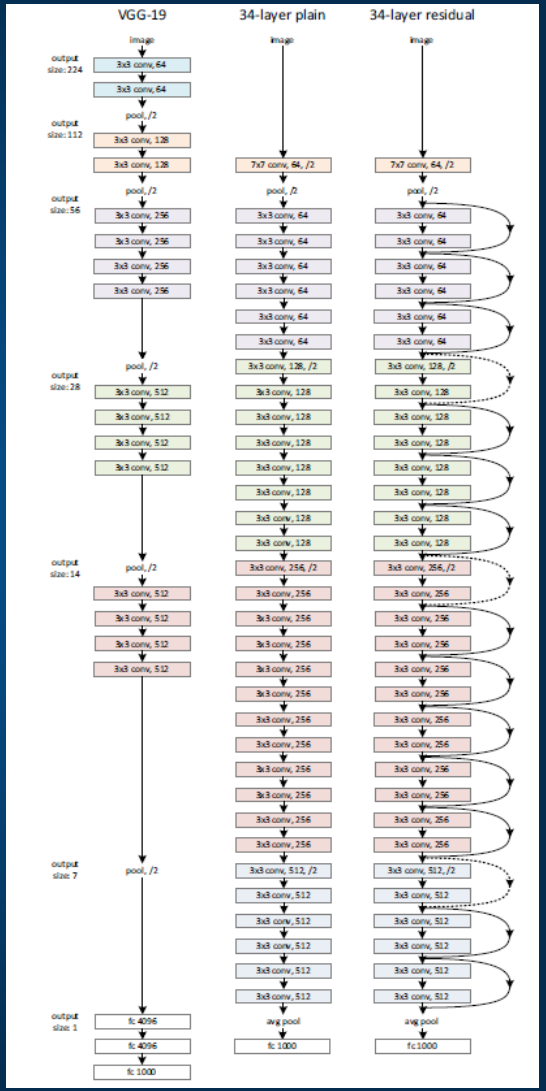
Dataset is CIFAR 10 or 100 or ImageNet



VGG: Oxford Group Network

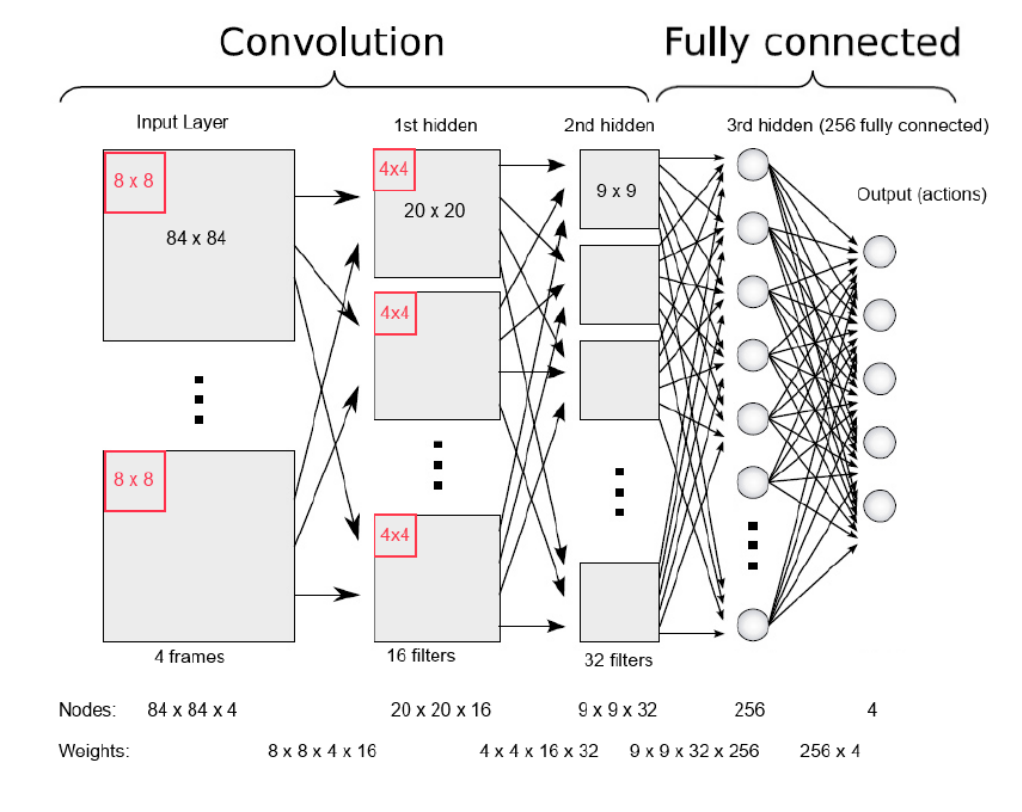
Paper: Karen Simonyan and Andrew Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition"

Dataset is CIFAR 10 or 100 or ImageNet



Deep Reinforcement Learning

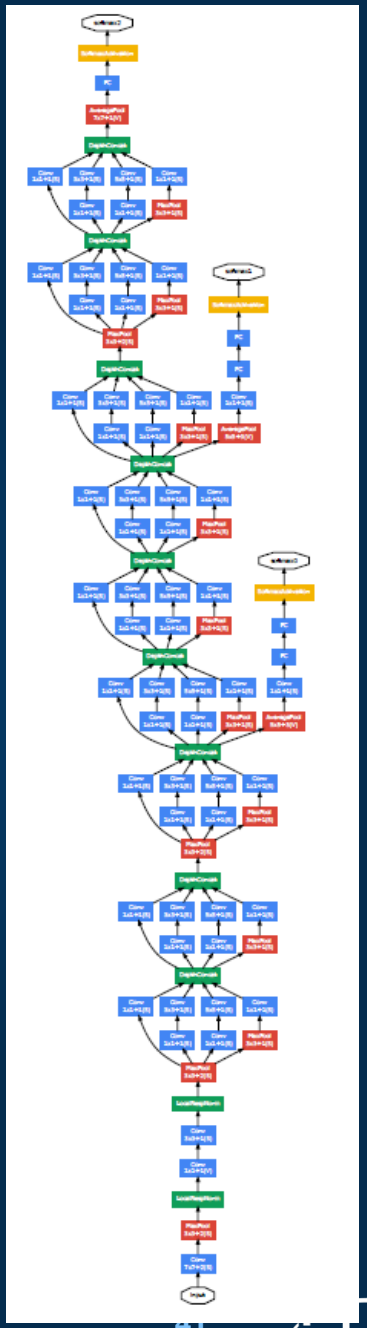
Paper: Volodymyr Minh et. al. "Playing Atari with Deep Reinforcement Learning".
Dataset: http://www.arcadelearningenvironment.org/



Inception/GooLeNet

Paper: Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. Going deeper with convolutions

Dataset: CIFAR 10 or 100 or ImageNet



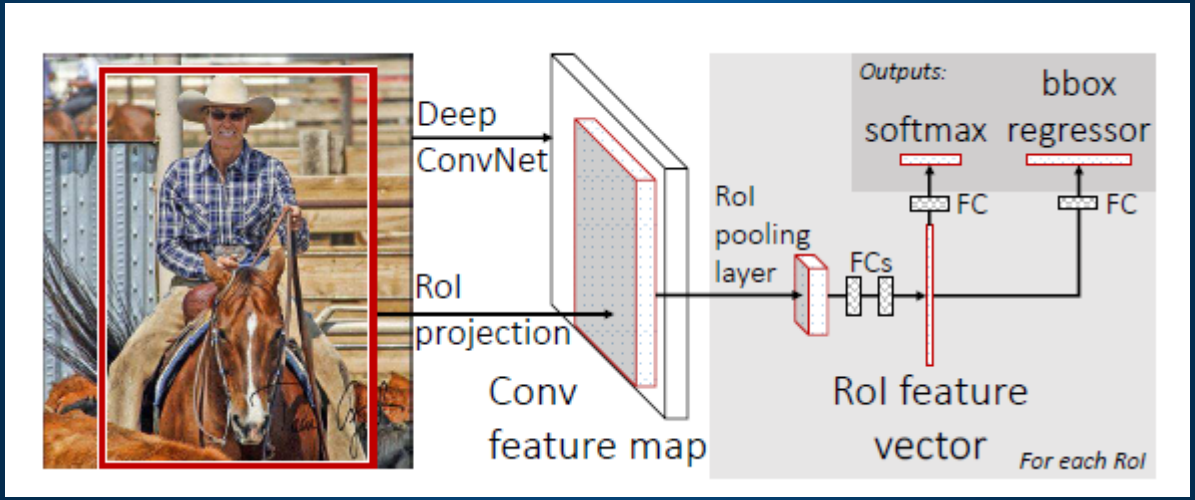
ICRI-CI Intel Collaborative Research Institute Computational Intelligence

intel

Fast R-CNN

Paper: Ross Girshick "Fast R-CNN"

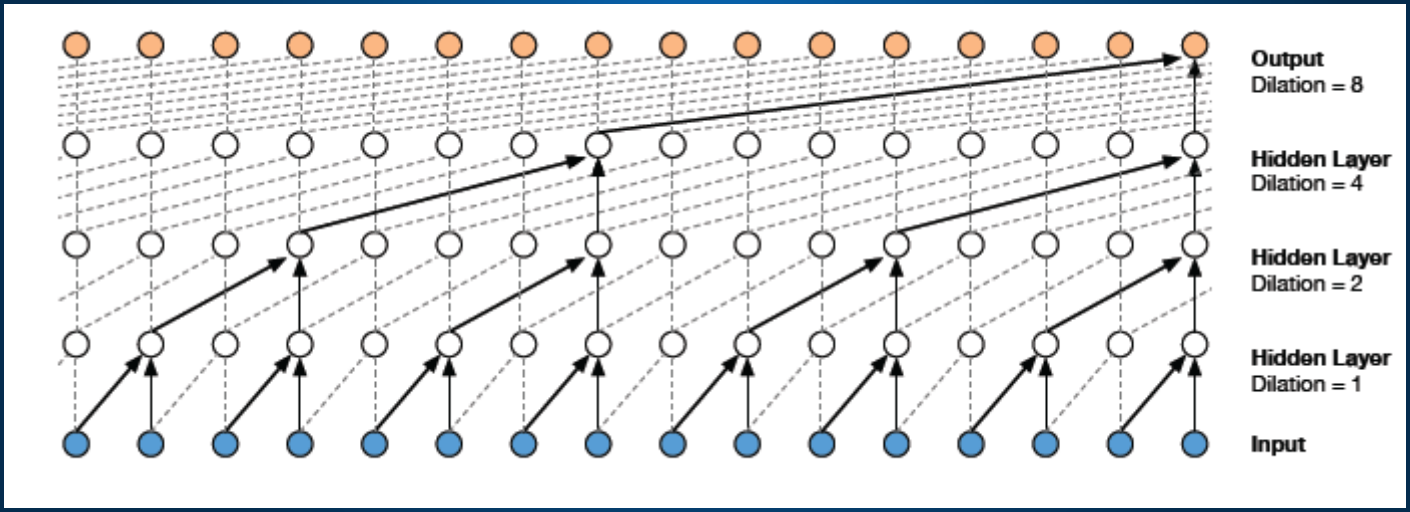
Datasets: Pascal VOC07



WaveNet

Paper: van den Oord et al: "WAVENET: A GENERATIVE MODEL FOR RAW AUDIO"

Dataset: VCTK (see paper)



Others

Object detection and localization

SSD: Single shot multibox detector
https://arxiv.org/abs/1512.02325
in Caffe here: https://github.com/weiliu89/caffe/tree/ssd

R-FCN: Object detection via region based fully convolutional networks:
https://arxiv.org/abs/1605.06409
Caffe code here: https://github.com/daijifeng001/caffe-rfcn

Semantic Segmentation:

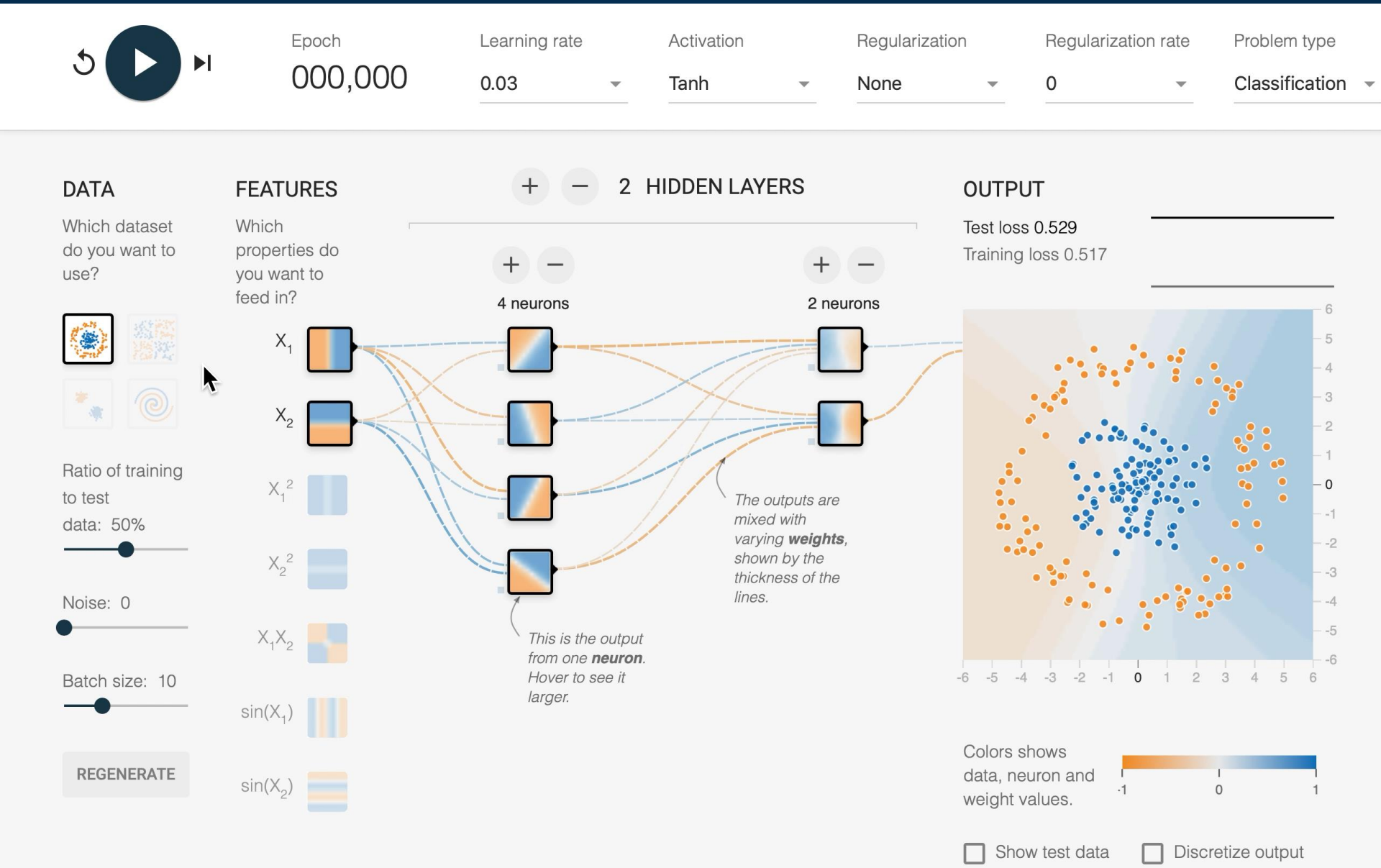
SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling
http://arxiv.org/abs/1505.07293
Code in caffe here: https://github.com/alexgkendall/caffe-segnet

Fully Convolutional Networks for Semantic Segmentation
https://arxiv.org/abs/1411.4038
In Caffe branch: https://github.com/BVLC/caffe

Try it yourself

for example,

playground.tensorflow.org


Epoch
000,000
Learning rate
0.03
Activation
Tanh
Regularization
None
Regularization rate
0
Problem type
Classification
DATA
Which dataset do you want to use?
Ratio of training to test data: 50%
Noise: 0
Batch size: 10
REGENERATE
FEATURES
Which properties do you want to feed in?
X1
X2
X1²
X2²
X1X2
sin(X1)
sin(X2)
2 HIDDEN LAYERS
4 neurons
2 neurons
The outputs are mixed with varying weights, shown by the thickness of the lines.
This is the output from one neuron. Hover to see it larger.
OUTPUT
Test loss 0.529
Training loss 0.517
Colors shows data, neuron and weight values.
Show test data
Discretize output