Deep Neural Networks

Brains Minds and Machines summer school 2016

Gemma Roig

Overview

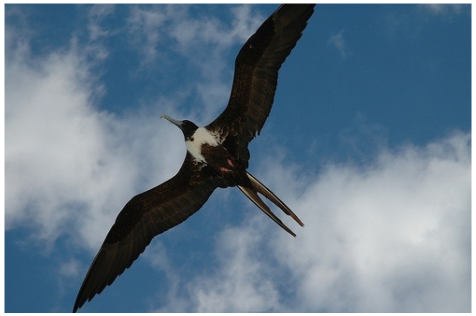
- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition
- Applications
- Limitations and Open Questions

Object Recognition

What is in the image?

Bike

Train



Bird

Reminder

We want the algorithms to **learn** to do object recognition given examples of the object category

Training phase: examples images are shown to the algorithm

Testing phase: labelling of images never shown before

There are different modalities of supervision (fully supervised, unsupervised, semi-supervised, etc.)

Overview

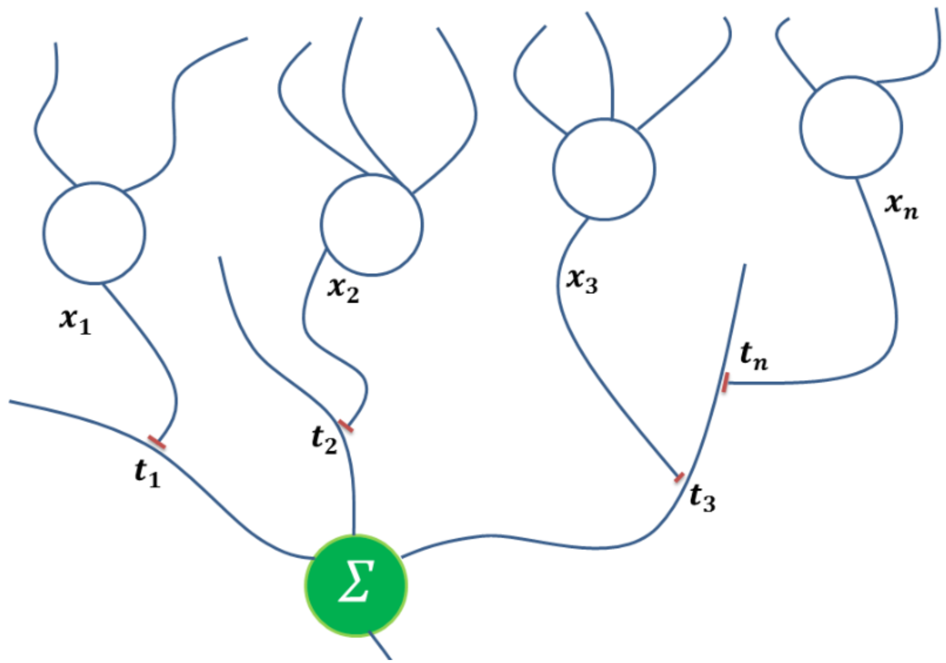
- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition
- Applications
- Limitations and Open Questions

Principles

Simplified neuroscience: a neuron computes a dot product between its inputs and the synaptic weights

$$< x, t > \longleftrightarrow$$

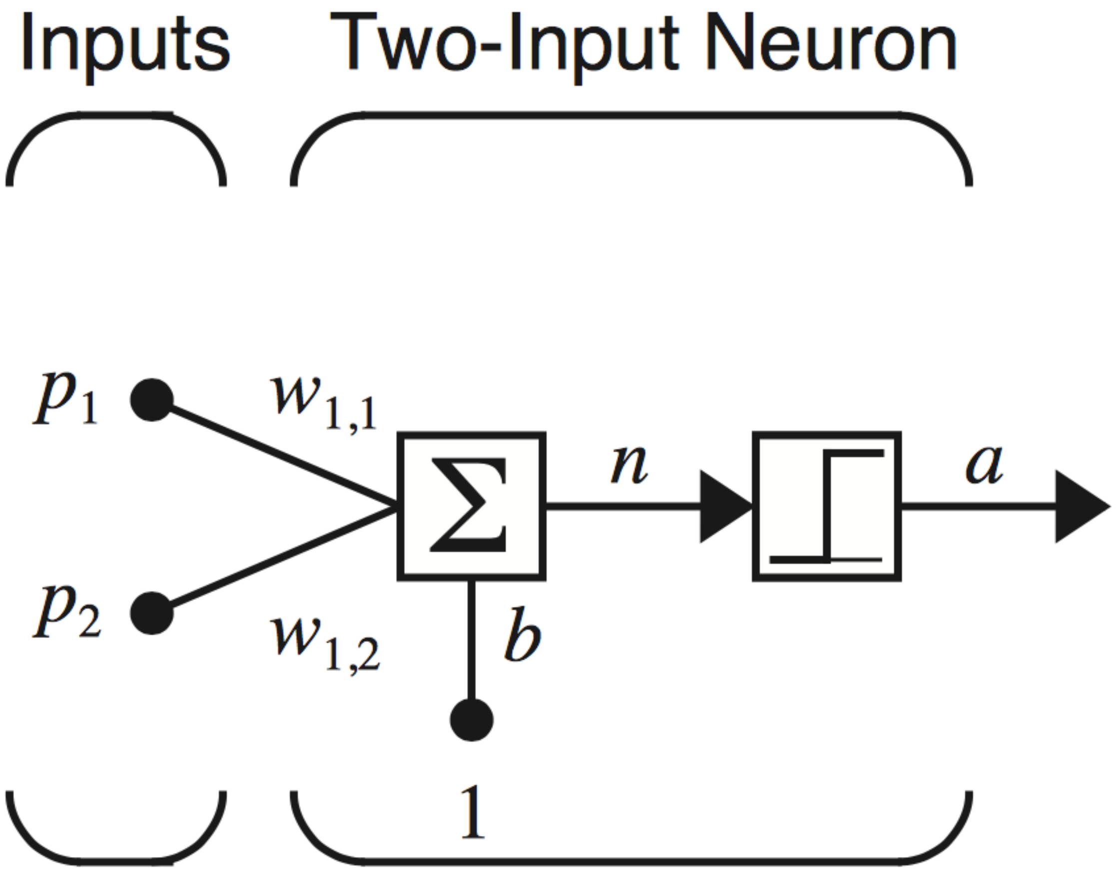
Neuroscience definition of dot product!



Perceptron

F. Rosenblatt 1957

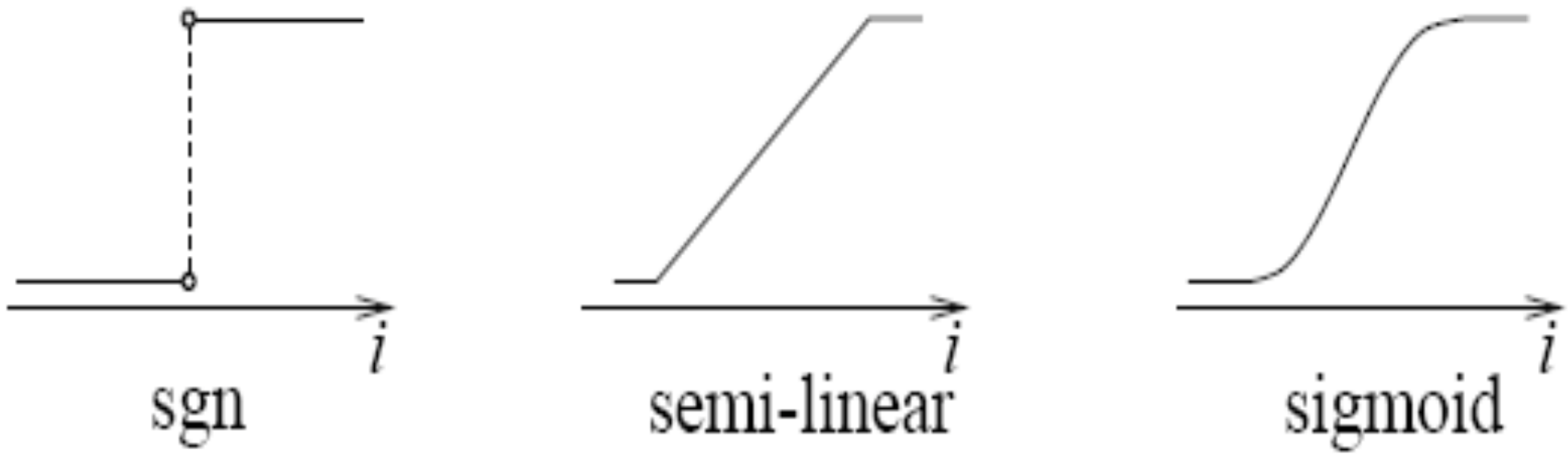
One layer NN



$$a = hardlim(\mathbf{Wp} + b)$$

Perceptron

Types of Nonlinearities



etc.

Learning

Gradient descend

$$\mathbf{w} \leftarrow \mathbf{w} + \mathbf{x}_i(y_i - y_i^*)$$

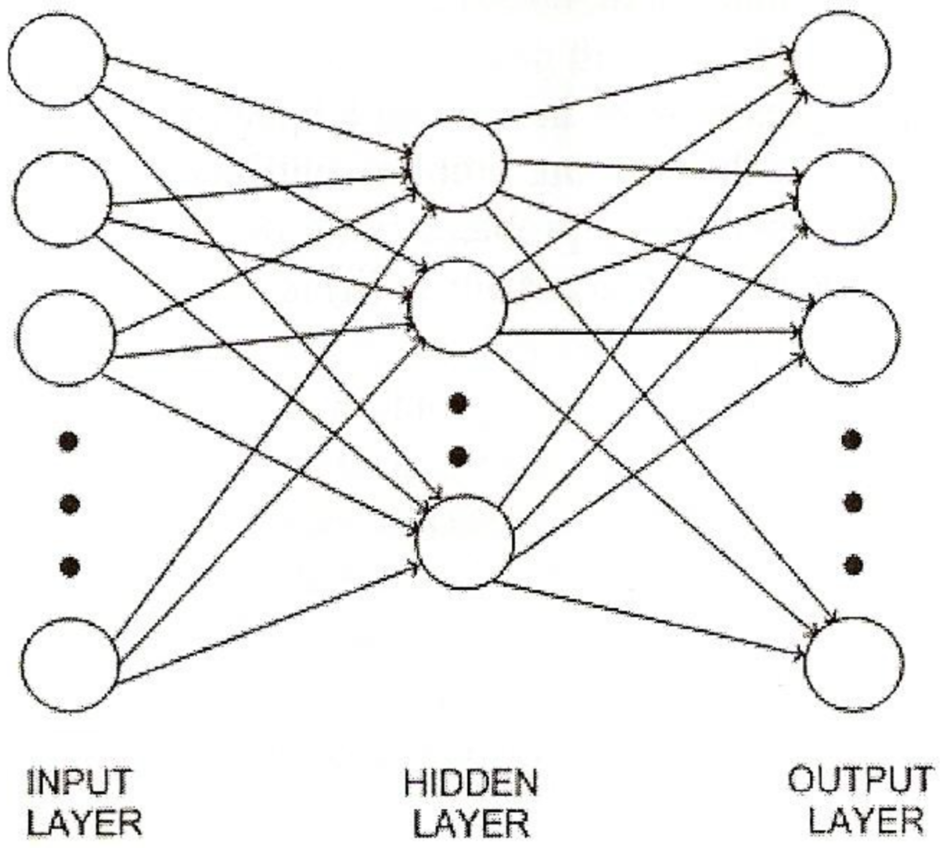
In case of linear separable data, the learning converges in a number of iterations that can be bounded by (R/γ)^2.

R is the norm of the largest input vector,

γ is the margin between the decision boundary and the closest data-case.

Multi-layer Perceptron

Rumelhart et al. 1986



and possibly many more layers

Back-propagation

Learning based on iterating between:

1. Propagation

    1.1. Forward pass through NN

    1.2 Backward pass using derivatives

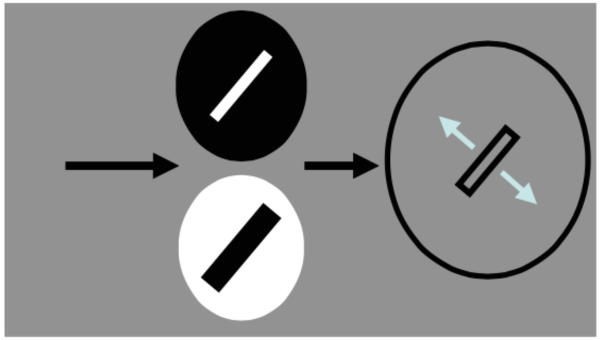
2. Weights updates

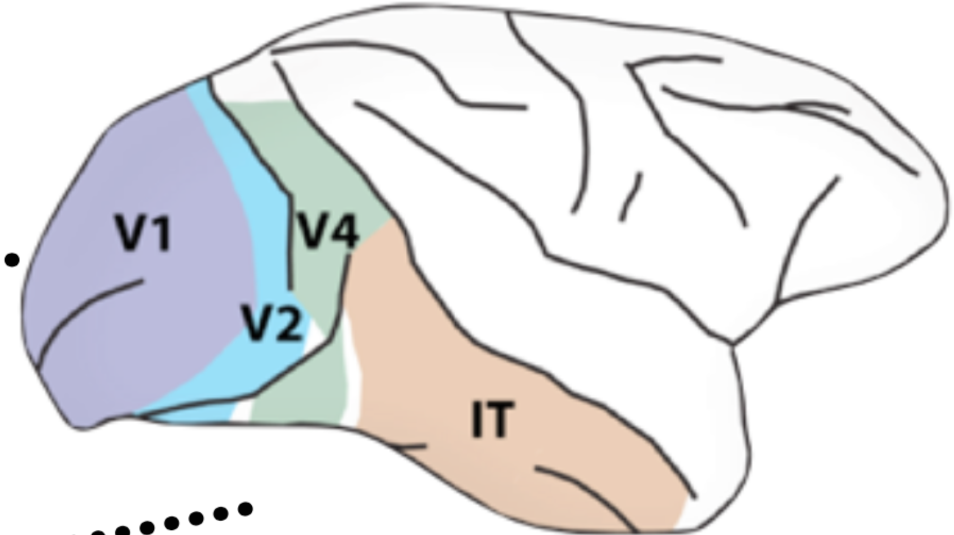
(gradient descend)

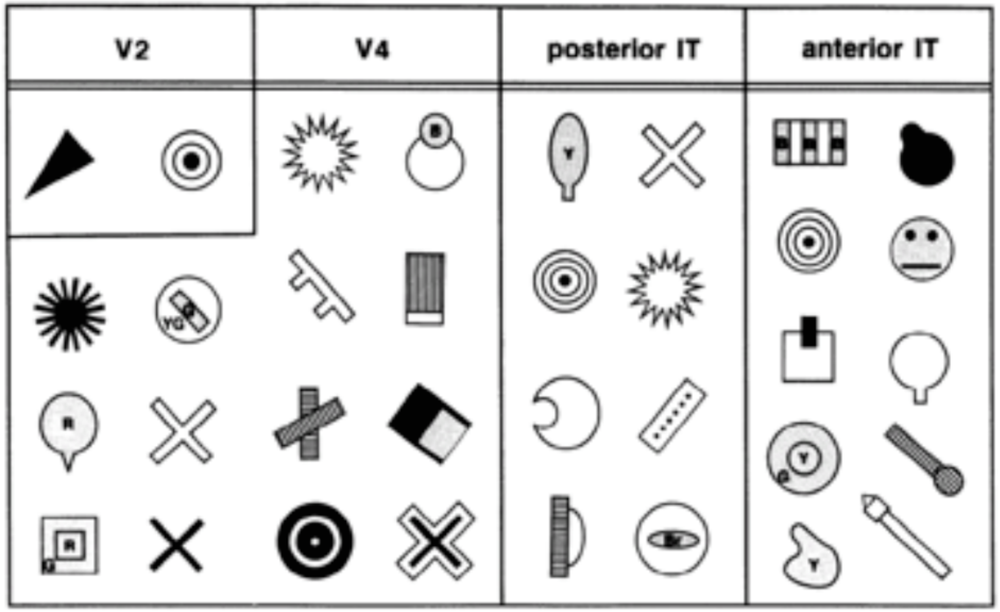
Overview

- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition
- Applications
- Limitations and Open Questions

The ventral stream







The ventral stream hierarchy: V1, V2, V4, IT
A gradual increase in the receptive field size, in the complexity of the preferred stimulus, in tolerance to position and scale changes

Kobatake & Tanaka, 1994

Hubel and Wiesel

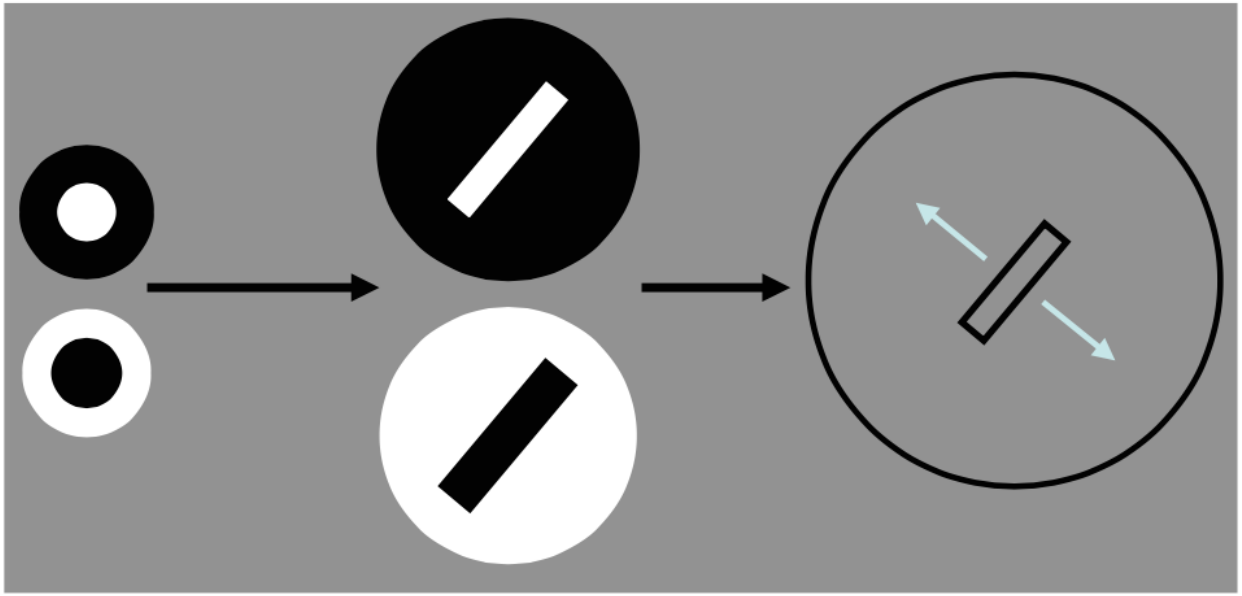
(1959) Nobel prize

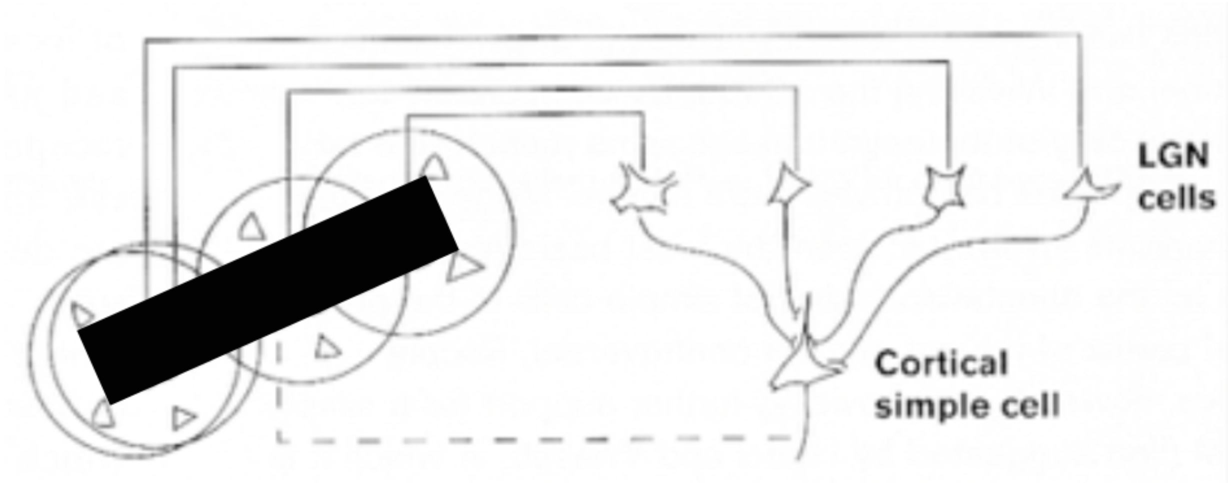
-> See Videos

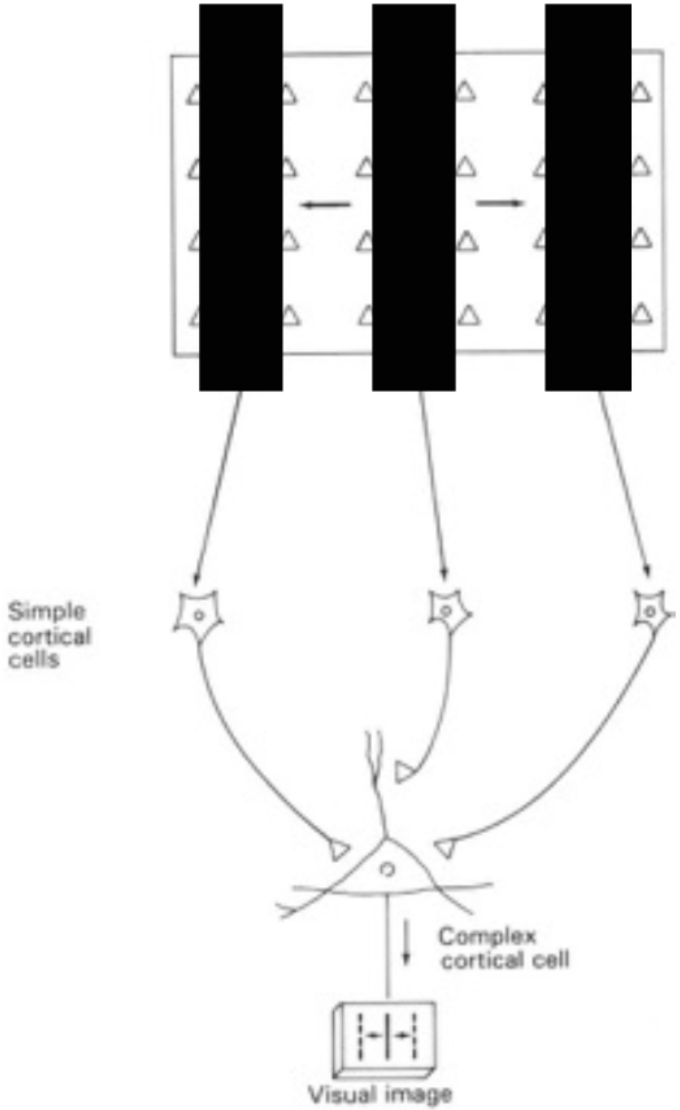
https://www.youtube.com/watch?v=IOHayh06LJ4

https://www.youtube.com/watch?v=jw6nBWo21Zk

LGN-type cells | Simple cells | Complex cells

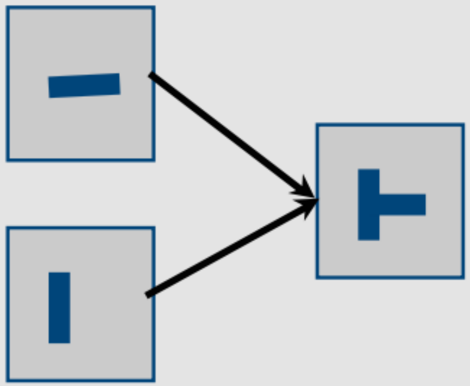
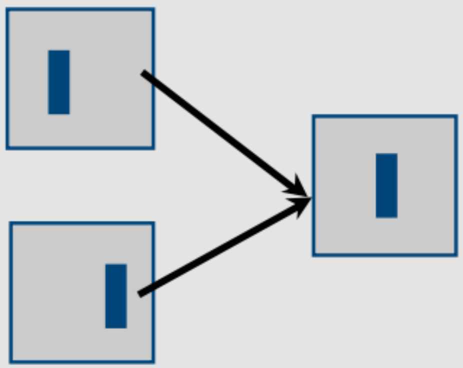






(Hubel & Wiesel 1959)

Simple and Complex Cells

Unit	Pooling	Computation
Simple		Selectivity / template matching
Complex		Invariance

Simple and Complex Cells

- Tuning operation (Gaussian-like, AND-like)

$$y = e^{-|x-w|^2}$$

$$or$$

$$y \sim \frac{x \bullet w}{|x|}$$

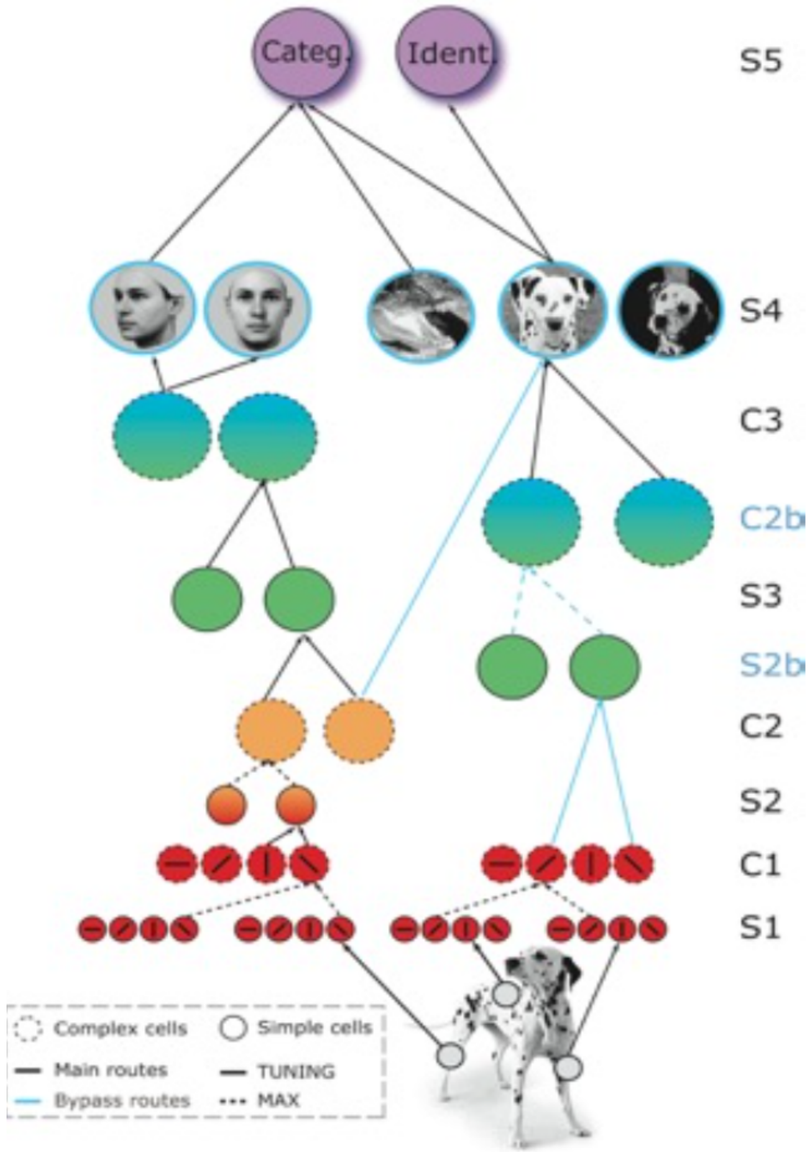
- Simple units

- Max-like operation (OR-like)

$$y = \max\{x1, x2, \dots\}$$

- Complex units

HMAX



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

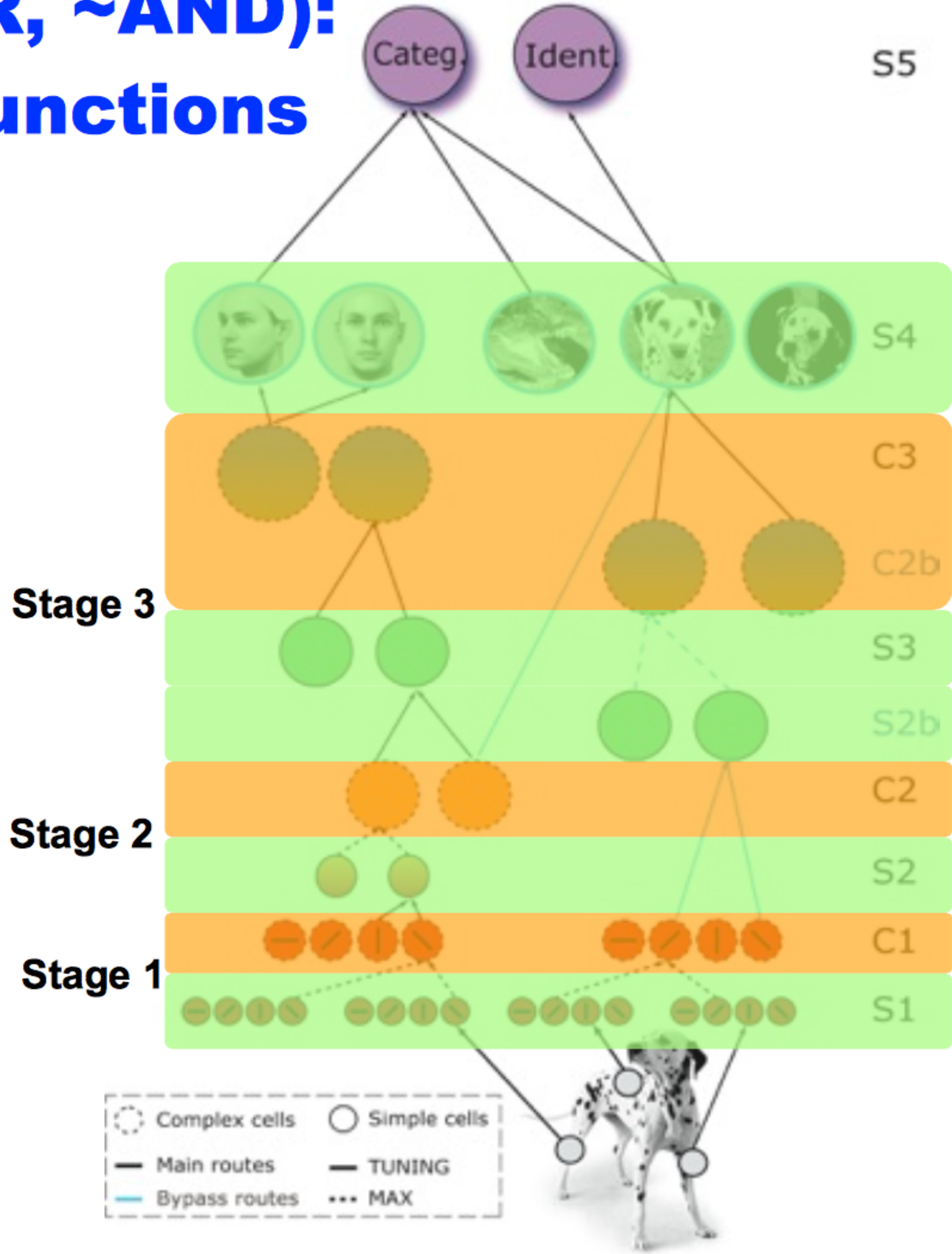
Two operations (~OR, ~AND): disjunctions of conjunctions

- Tuning operation (Gaussian-like, AND-like) $y = e^{-|x-w|^2}$ *or* $y \sim \dfrac{x \bullet w}{|x|}$
- Simple units

- Max-like operation (OR-like)

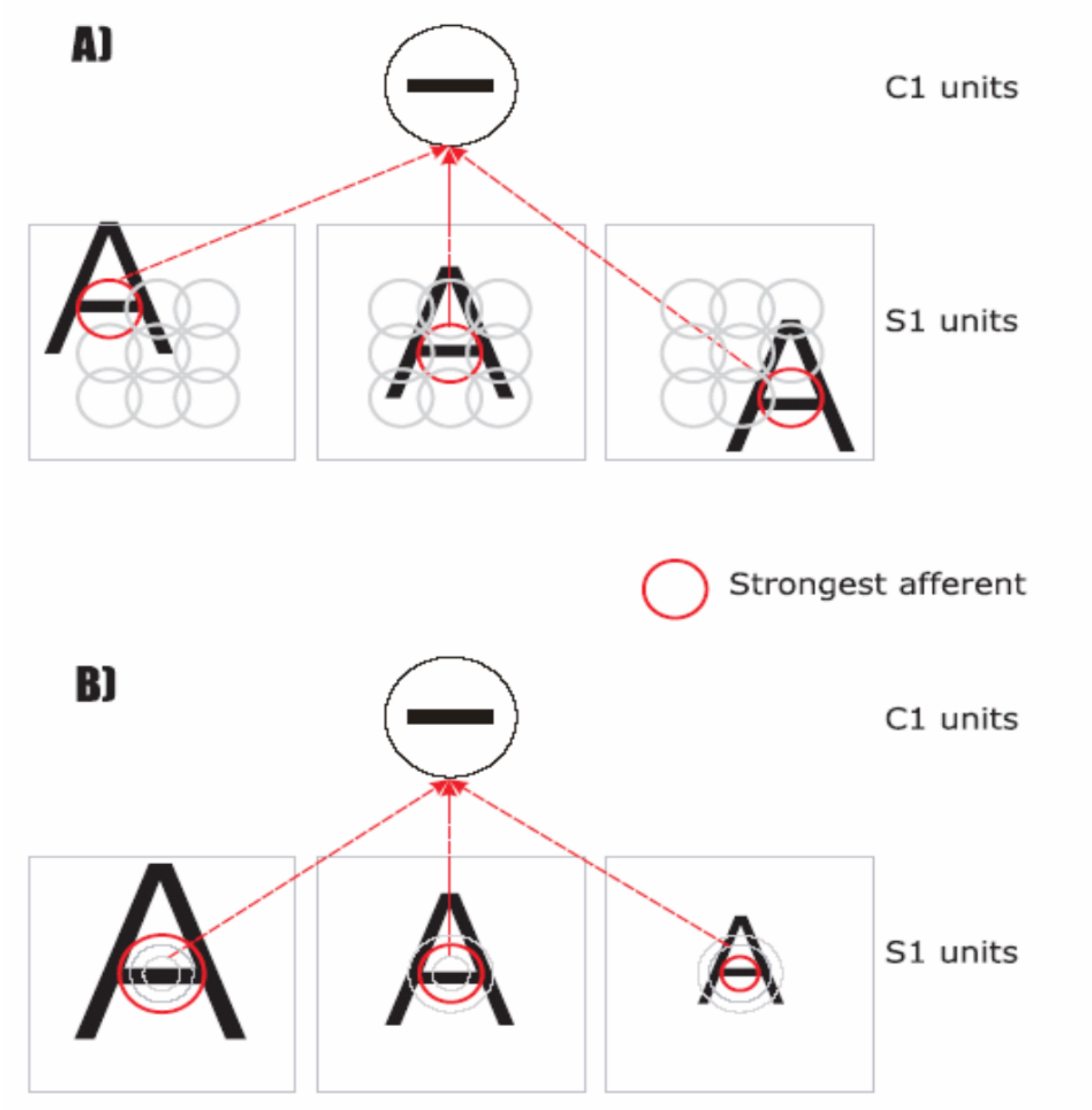
$$y = \max\{x1, x2, \ldots\}$$

- Complex units



Each operation ~microcircuits of ~100 neurons

Invariance

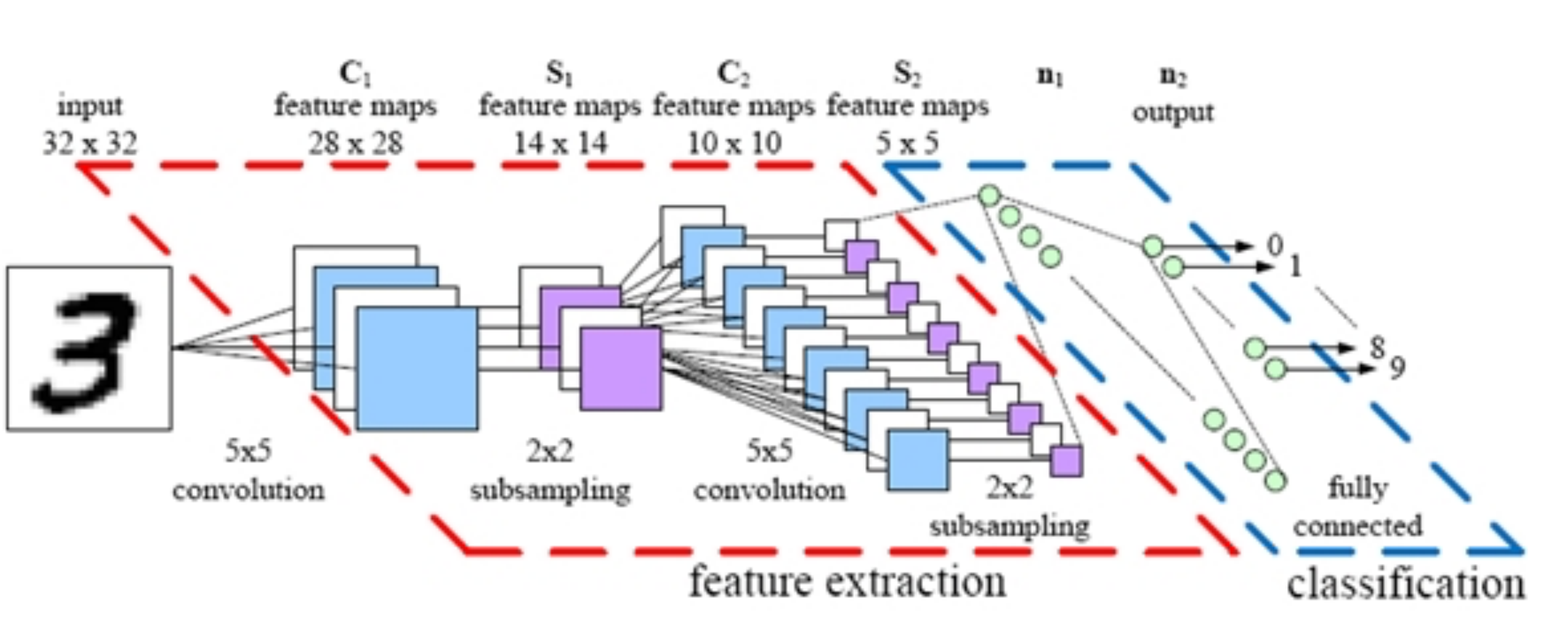


Serre, T., and Riesenhuber, M. (2004)

Overview

- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition
- Applications
- Limitations and Open Questions

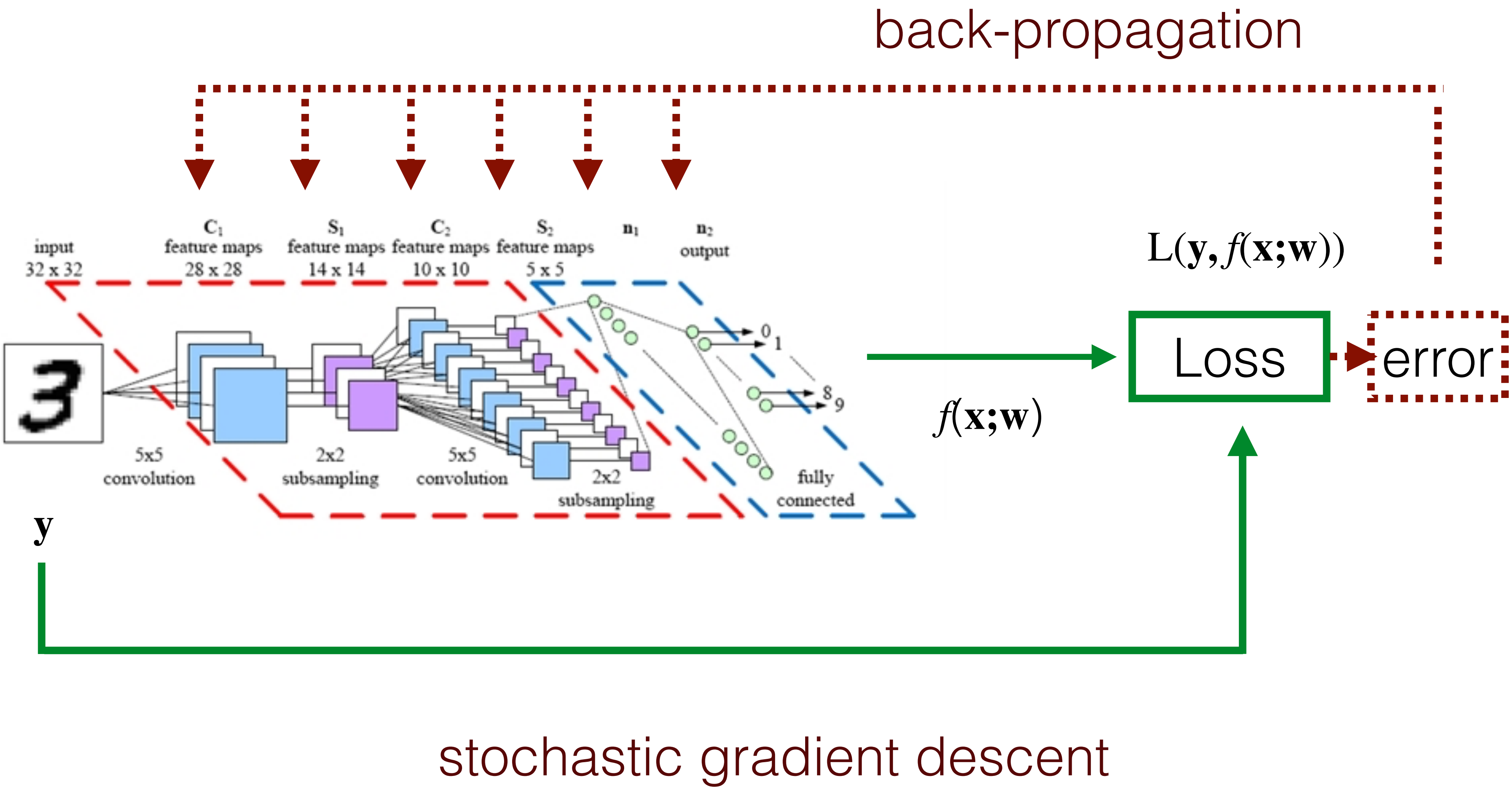
Convolutional Neural Networks



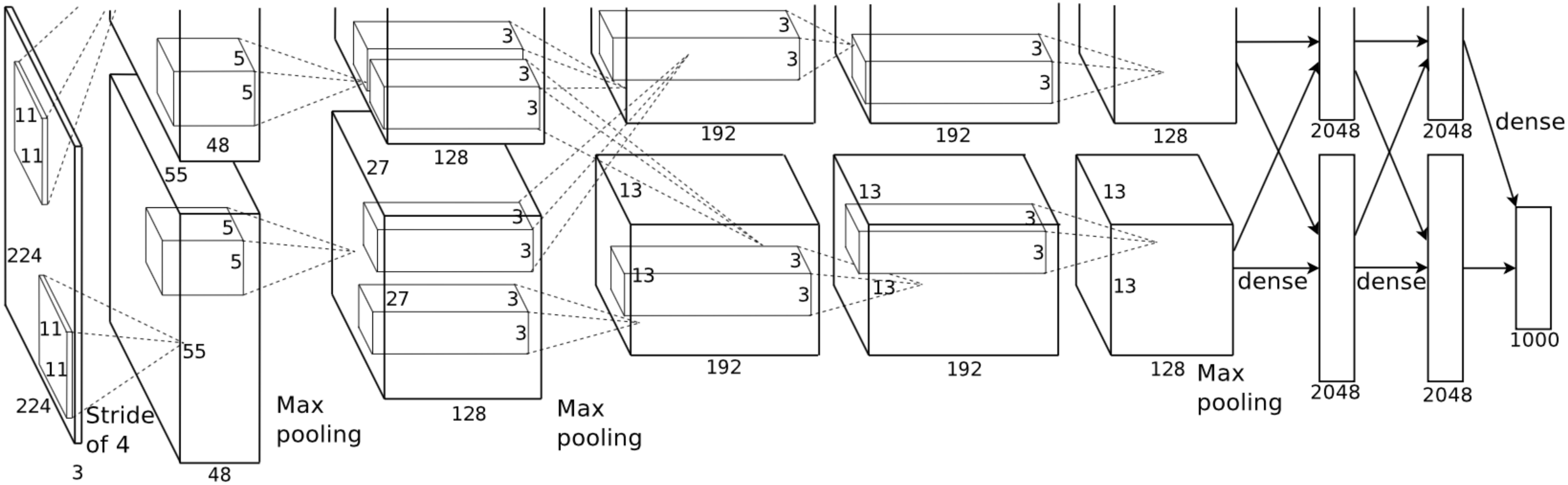
Emphasis on the convolutional assumption

LeCun et al. 98

Learning


back-propagation
input
32 x 32
C_1
feature maps
28 x 28
S_1
feature maps
14 x 14
C_2
feature maps
10 x 10
S_2
feature maps
5 x 5
n_1
n_2
output
5x5
convolution
2x2
subsampling
5x5
convolution
2x2
subsampling
fully
connected
0
1
8
9
$f(\mathbf{x};\mathbf{w})$
$L(\mathbf{y}, f(\mathbf{x};\mathbf{w}))$
Loss
error
$\mathbf{y}$
stochastic gradient descent

Deep CNN (2012)



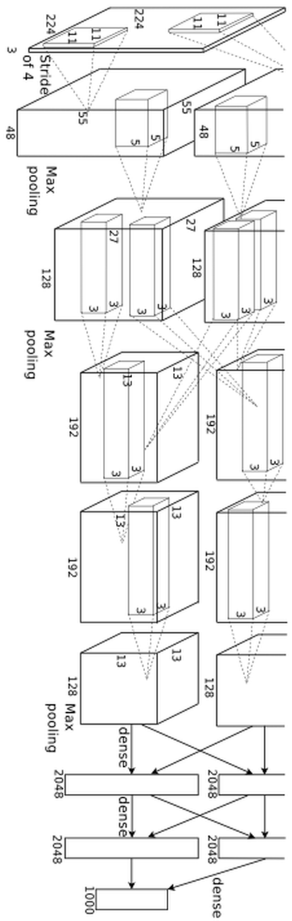
Learned with back propagation on GPUs (7 days)

Techniques to avoid overfitting

ImageNet dataset (1 million labeled images available)

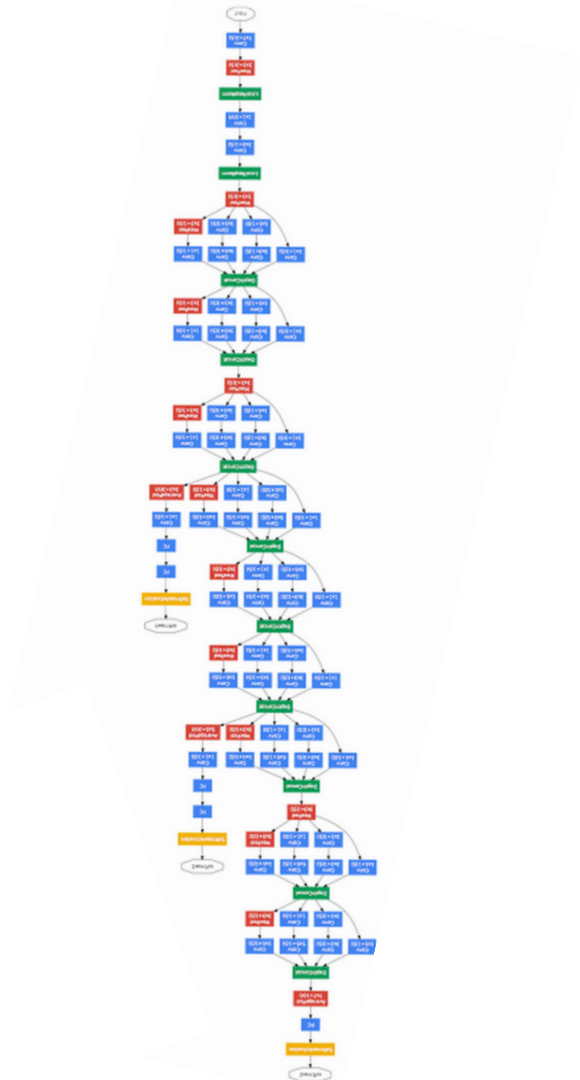
Krizhevsky et al. 12

Object classification



AlexNet 12

VGG 14



GoogLeNet 14

Avoid Overfitting

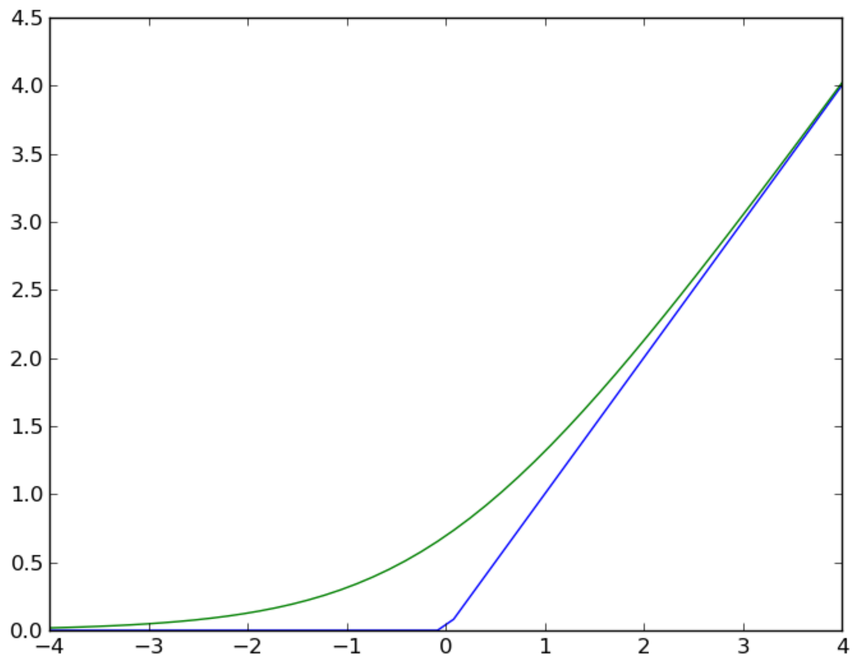
- Architecture of the network as prior:
  - convolutions
  - ReLU

- Use data augmentation in the training
  - affine transformations

- Dropout

Rectified Linear Unit

ReLU (blue line)



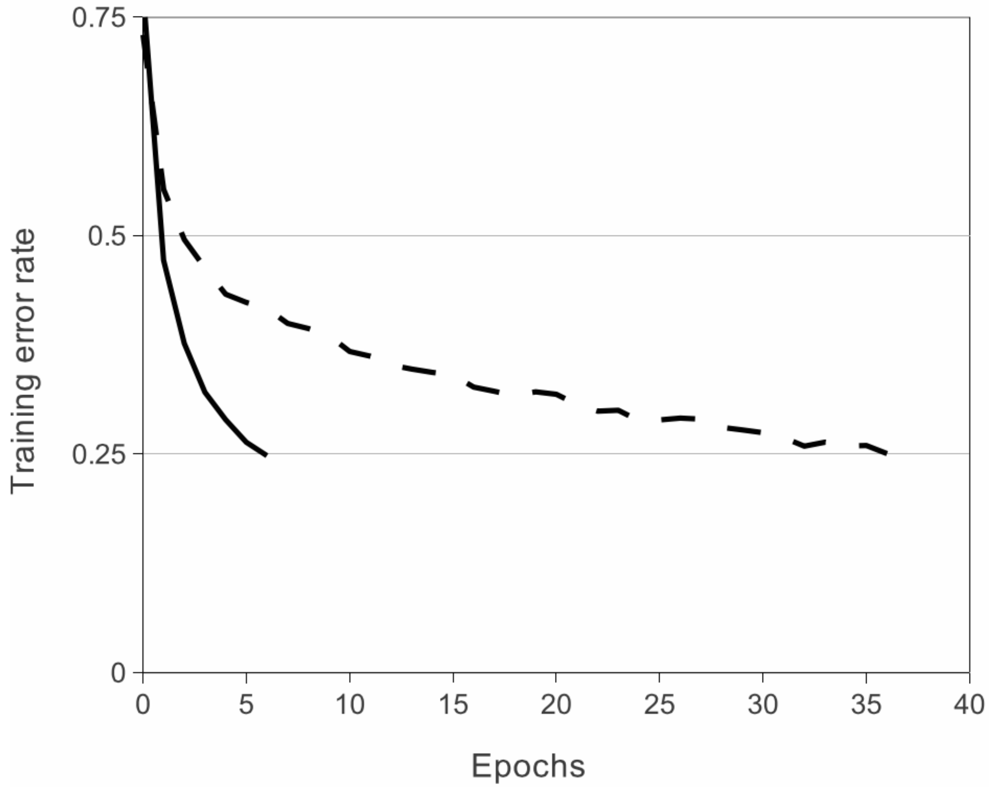


Figure 1: A four-layer convolutional neural network with ReLUs **(solid line)** reaches a 25% training error rate on CIFAR-10 six times faster than an equivalent network with tanh neurons **(dashed line)**. The learning rates for each net-

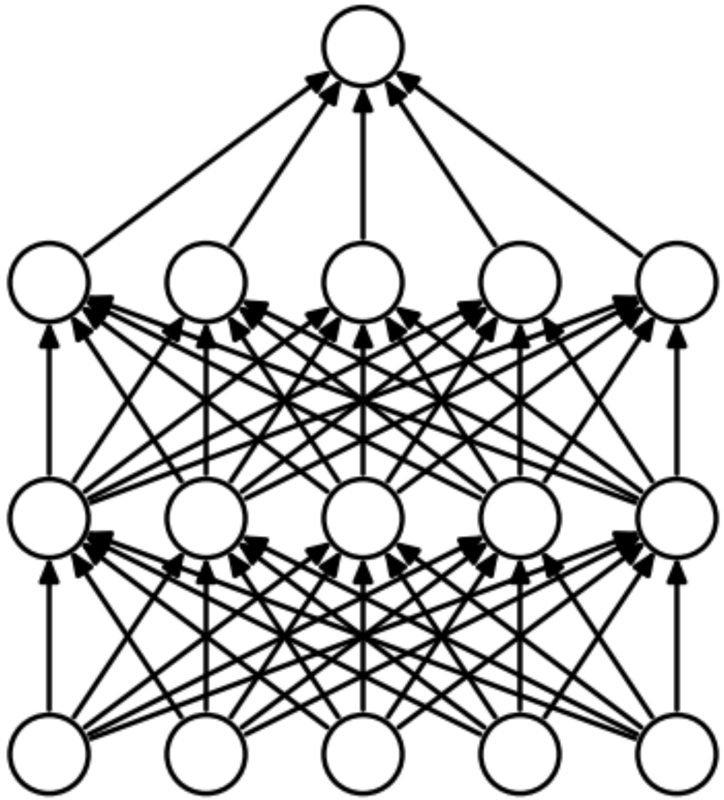
Krizhevsky et al. 12

Avoid Overfitting

- Dropout

training phase:
remove stochastically hidden units

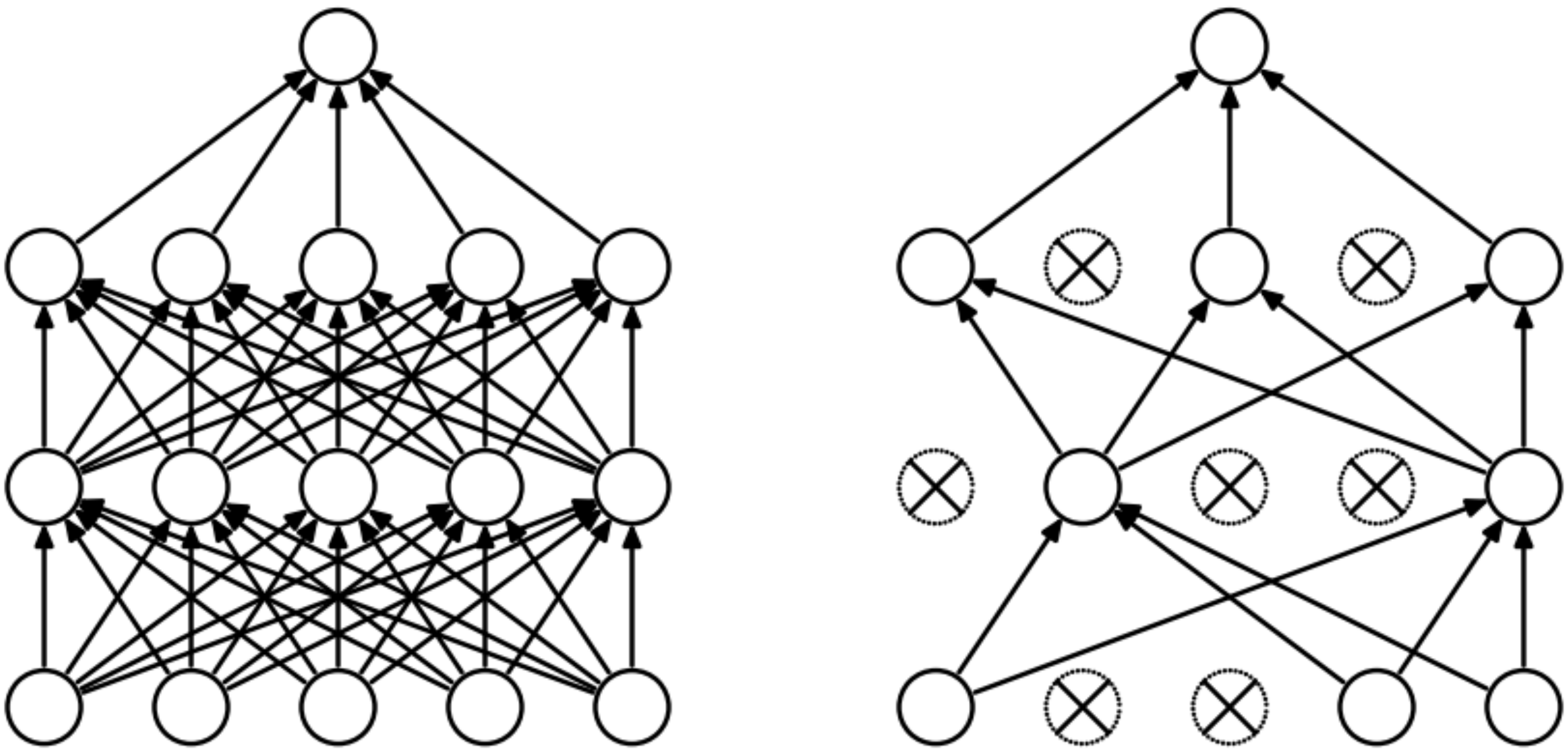
- hidden units set to 0 with a probability (0.5) (changes stochastically)
- hidden units can not co-adapt to other hidden units



(a) Standard Neural Net

Avoid Overfitting

- Dropout



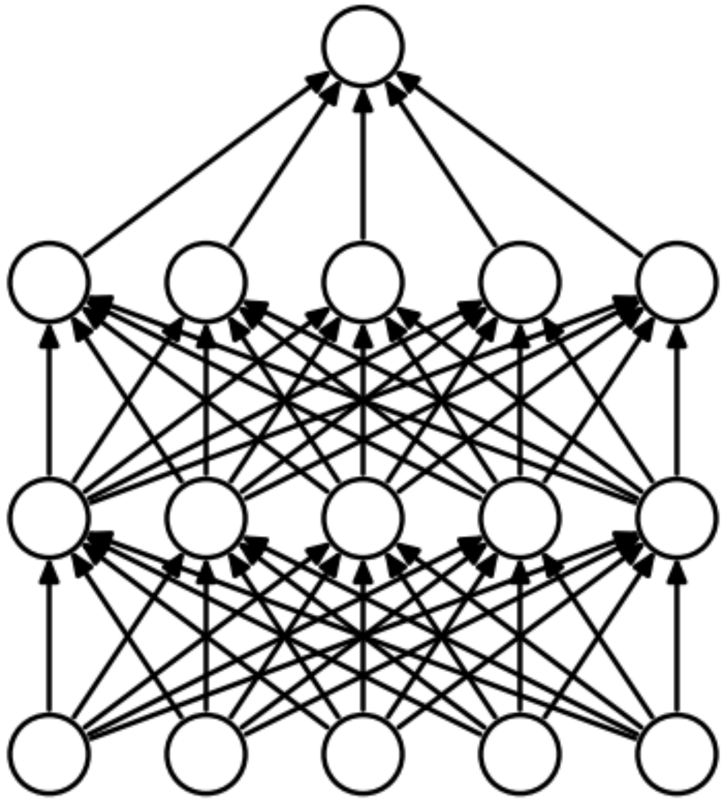
(a) Standard Neural Net (b) After applying dropout.

Avoid Overfitting

- Dropout

testing phase:
all hidden units used

- multiply hidden layers by the dropout probability (0.5) (not stochastic)
- better generalization



(a) Standard Neural Net

Amazing Results

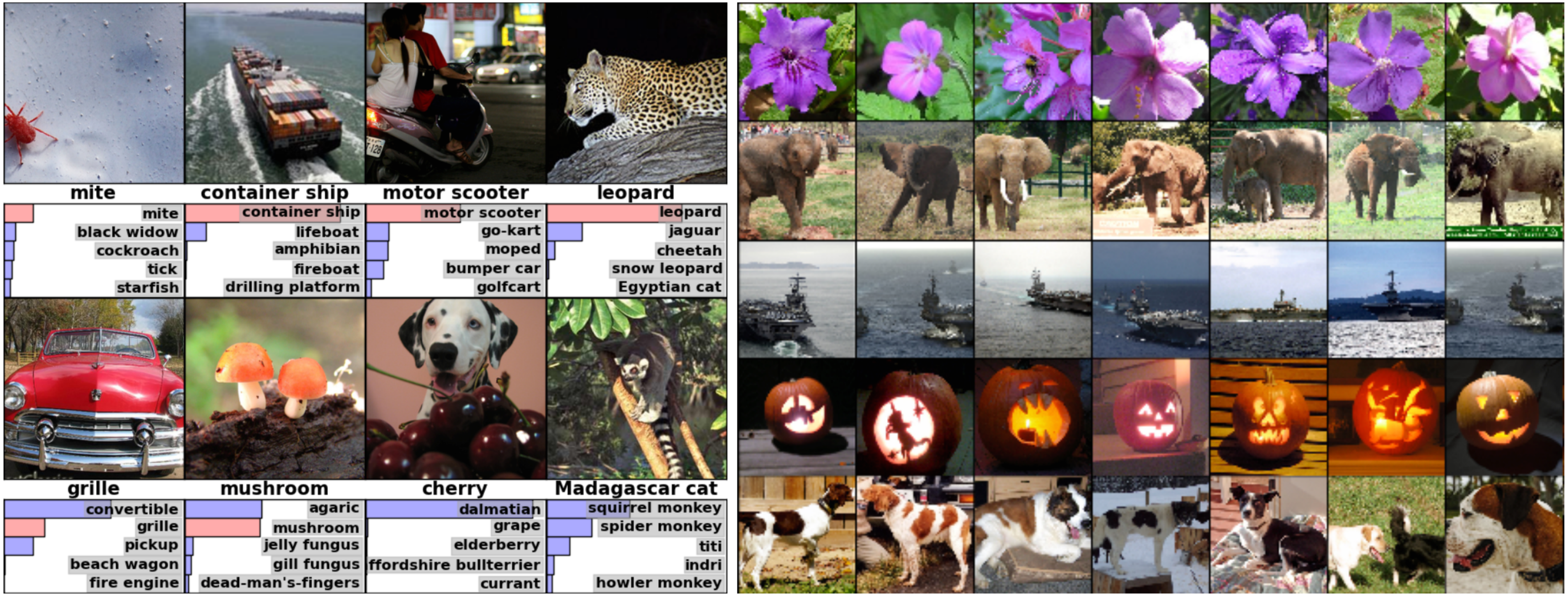
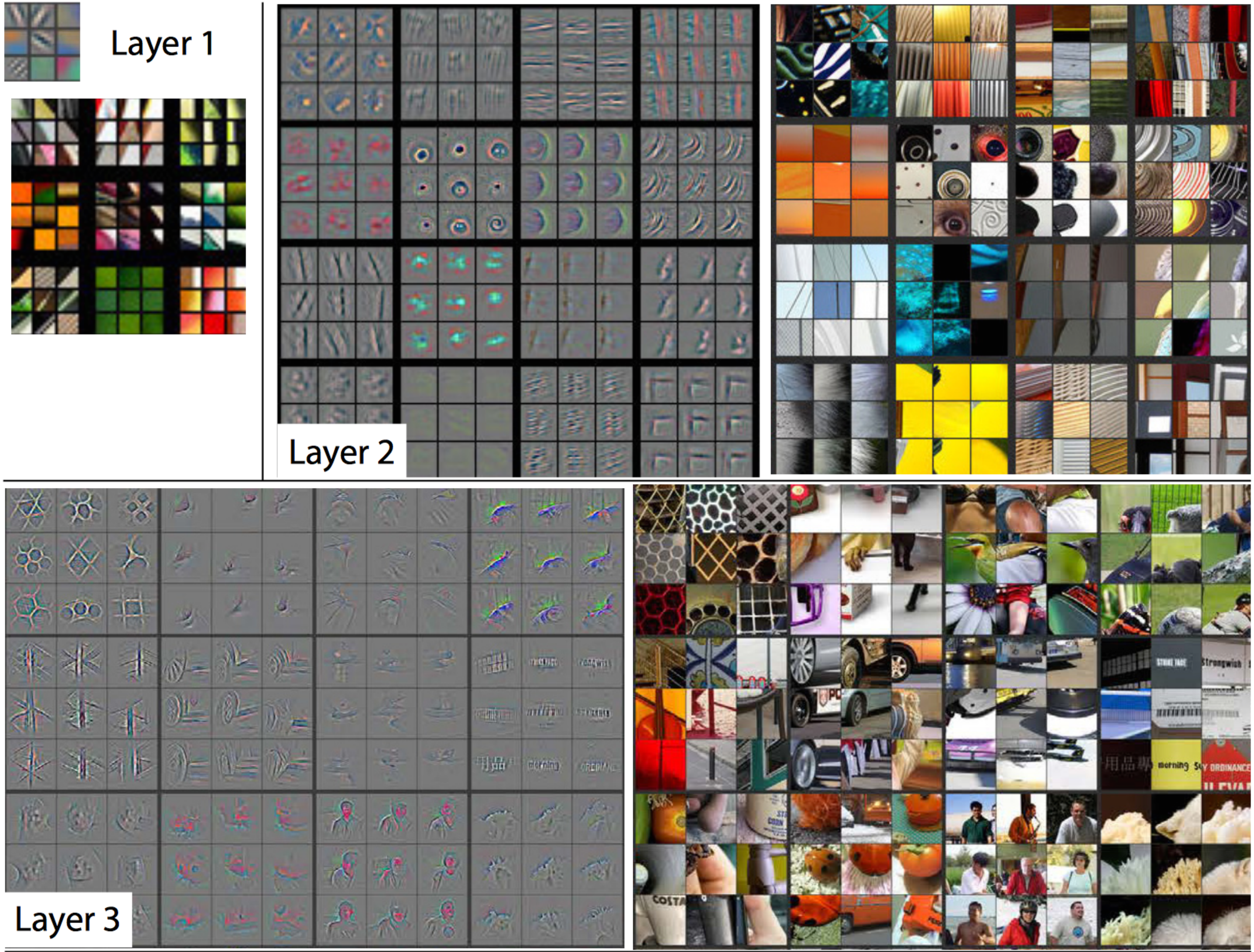
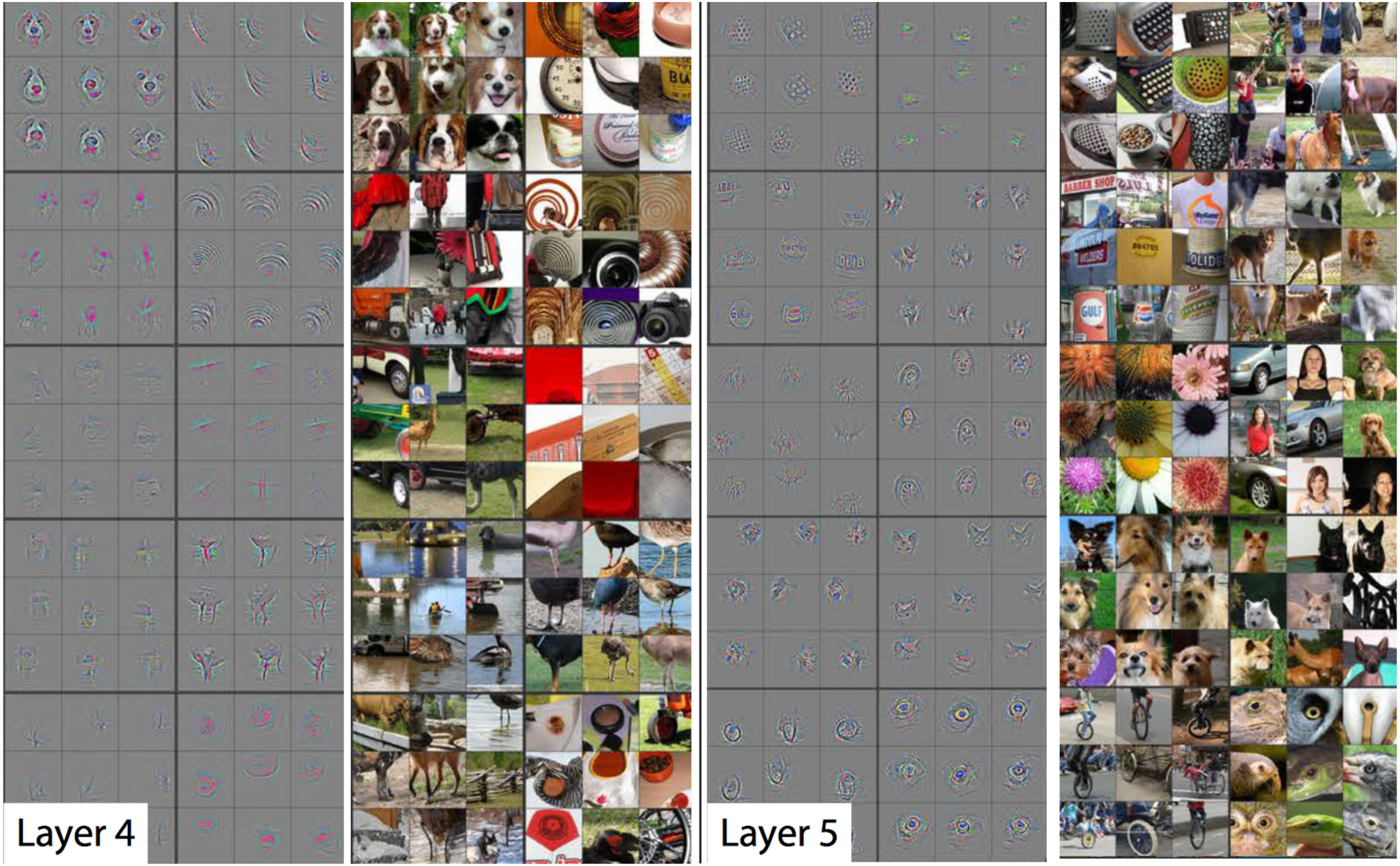


Figure 4: **(Left)** Eight ILSVRC-2010 test images and the five labels considered most probable by our model. The correct label is written under each image, and the probability assigned to the correct label is also shown with a red bar (if it happens to be in the top 5). **(Right)** Five ILSVRC-2010 test images in the first column. The remaining columns show the six training images that produce feature vectors in the last hidden layer with the smallest Euclidean distance from the feature vector for the test image.

Krizhevsky et al. 12

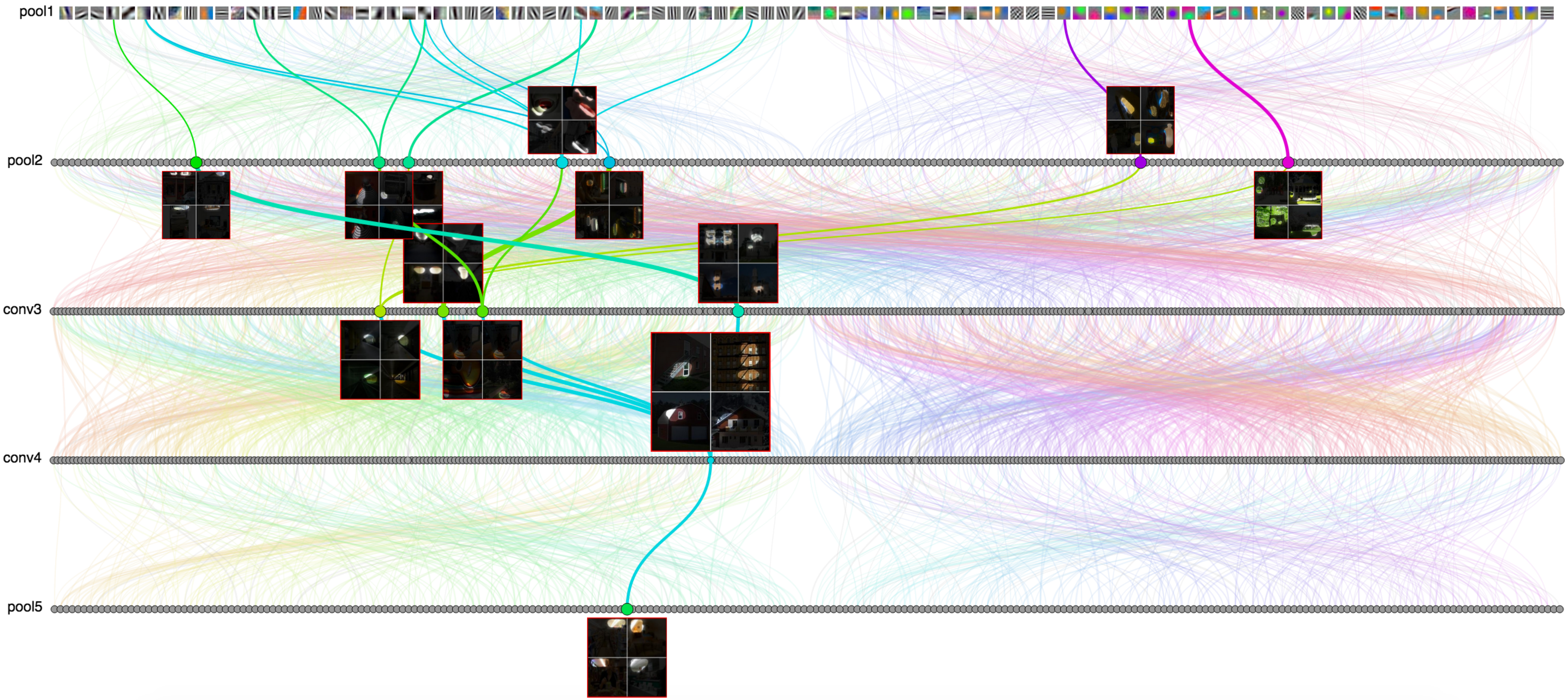
Visualization
of learned filters


Layer 1
Layer 2
Layer 3



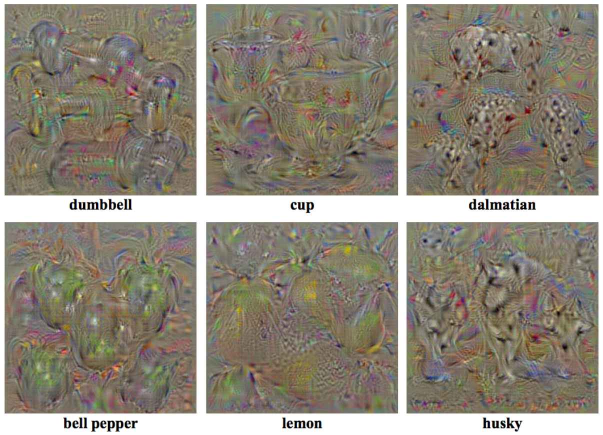
Zeiler and Fergus 13

Visualization of learned filters



http://people.csail.mit.edu/torralba/research/drawCNN/drawNet.html

Visualization of the DNN visual structure

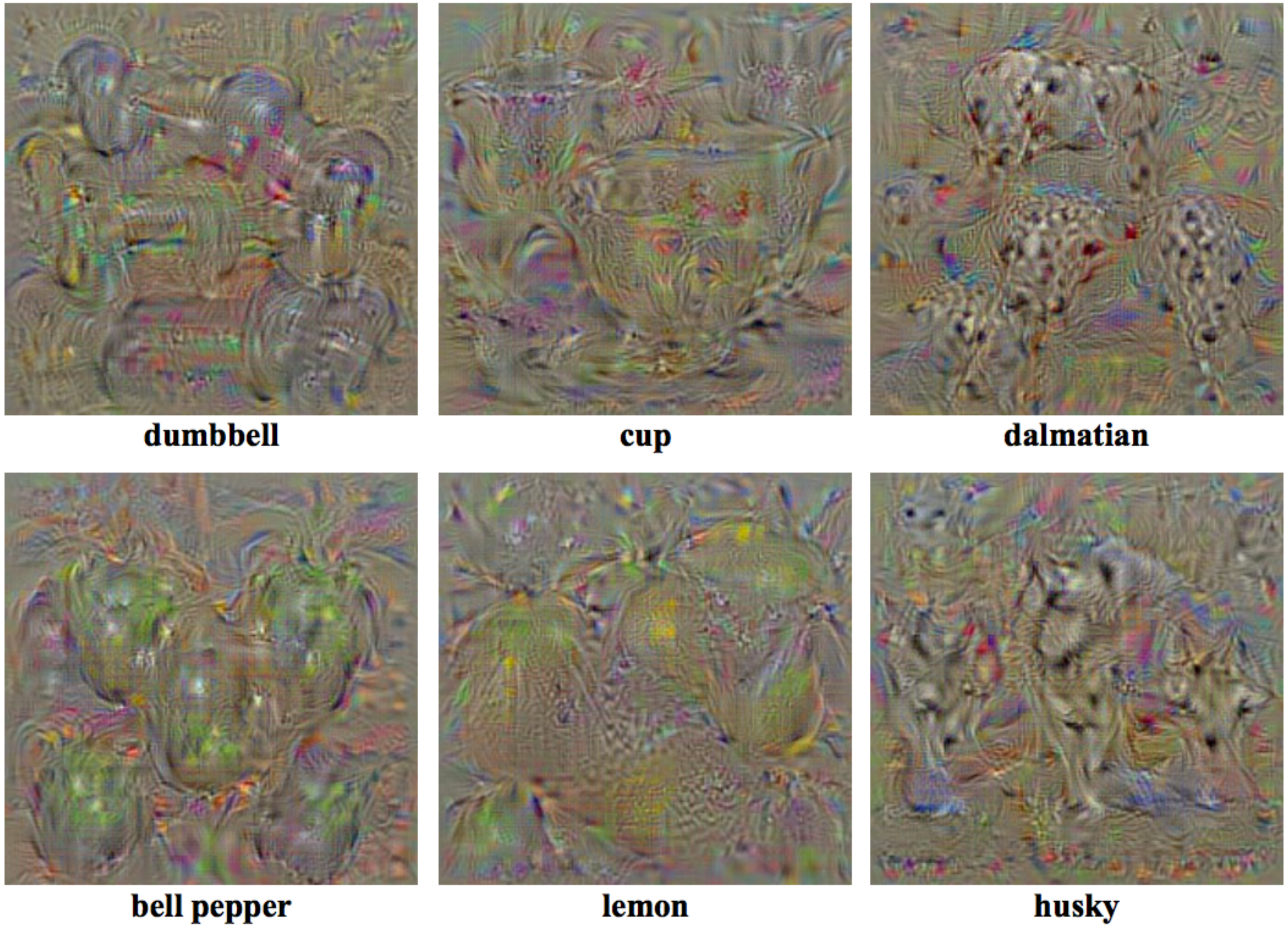


$$\arg\max_I S_c(I) - \lambda \|I\|_2^2$$

- Invert a CNN by finding the stimulus that maximizes the output of a class.

Simonyan et al. 2014

Visualization of the DNN visual structure



Simonyan et al. 2014

Invariance

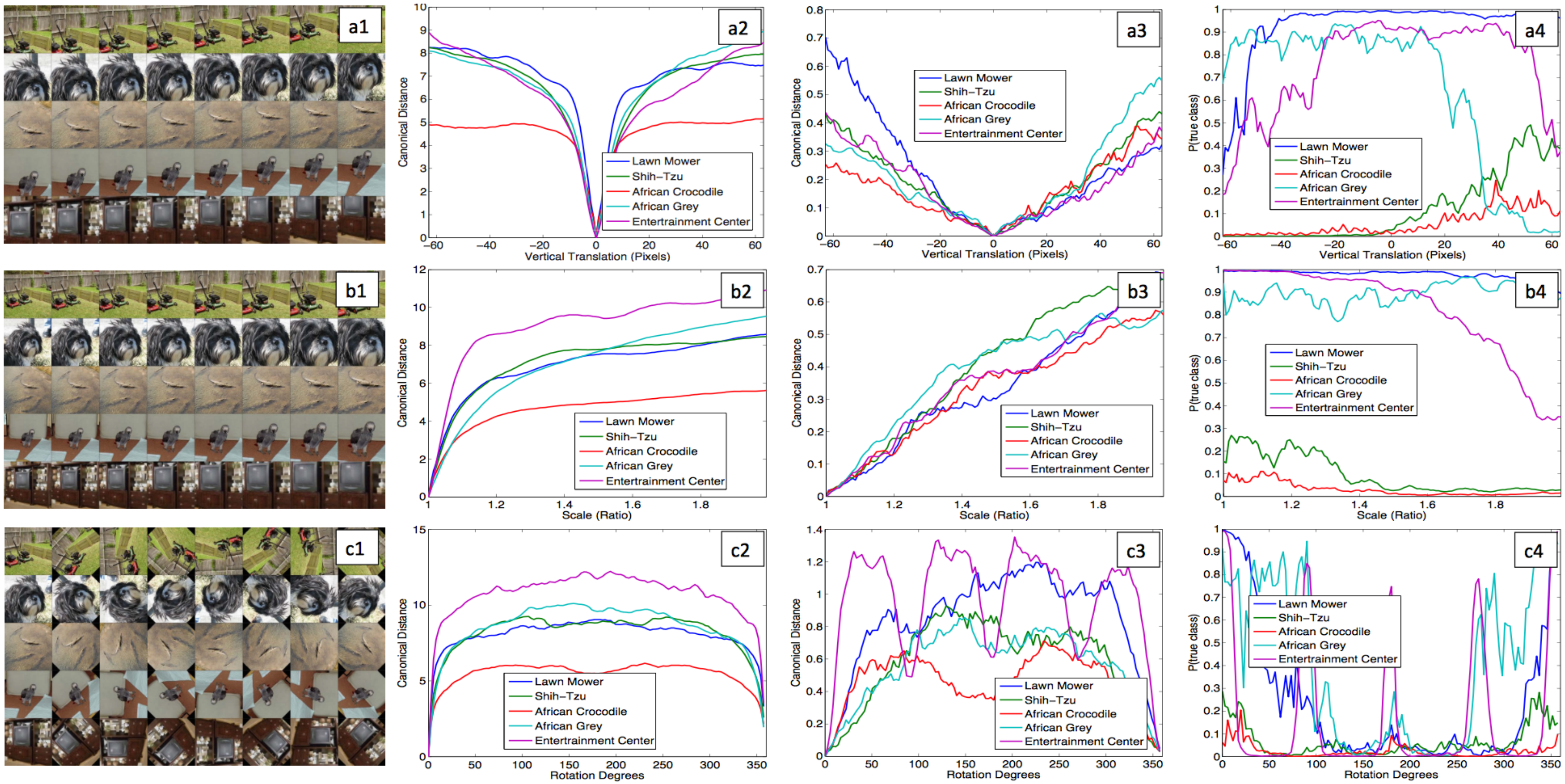


Figure 5. Analysis of vertical translation, scale, and rotation invariance within the model (rows a-c respectively). Col 1: 5 example images undergoing the transformations. Col 2 & 3: Euclidean distance between feature vectors from the original and transformed images in layers 1 and 7 respectively. Col 4: the probability of the true label for each image, as the image is transformed.

Zeiler and Fergus 13

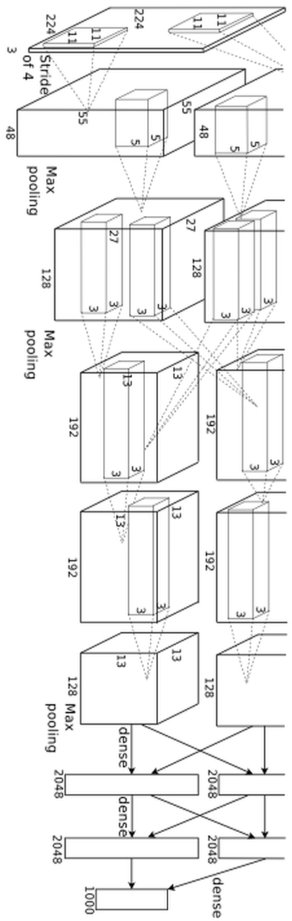
Overview

- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition
- Applications
- Limitations and Open Questions

Applications

- Use a pre-trained CNN as a feature extractor
- Fine-tune on limited data
- Train from scratch on big data

Object classification



AlexNet 12

VGG 14

GoogLeNet 14

Applications

- **Use a pre-trained CNN as a feature extractor**
- Fine-tune on limited data
- Train from scratch on big data

Object Detection

Rich feature hierarchies for accurate object detection

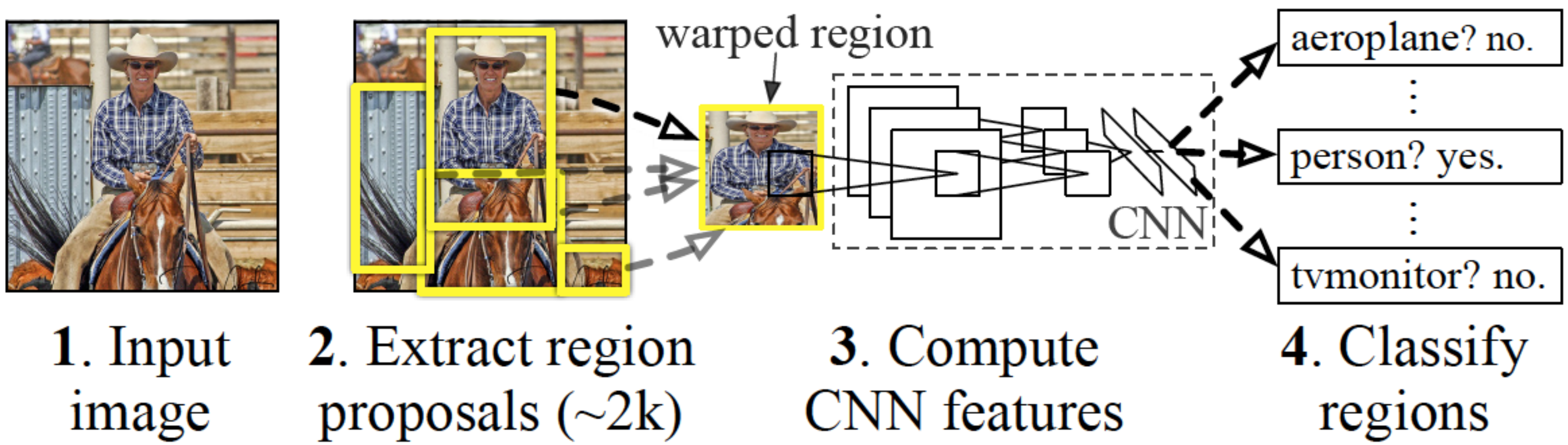
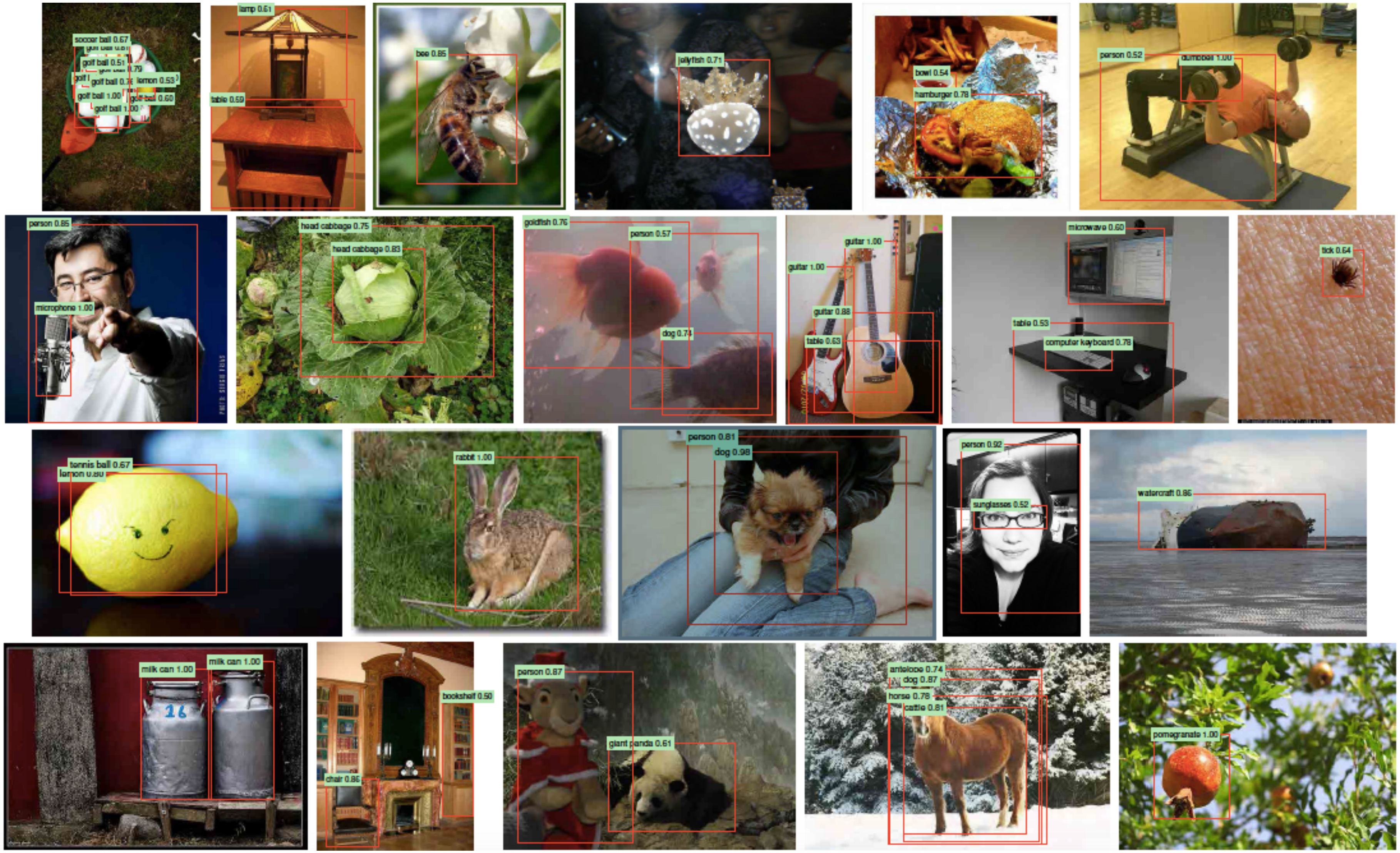


Figure 1: Object detection system overview. Our system (1) takes an input image, (2) extracts around 2000 bottom-up region proposals, (3) computes features for each proposal using a large convolutional neural network (CNN), and then (4) classifies each region using class-specific linear SVMs. R-CNN achieves a mean

Girshick et al. 14

Object Detection



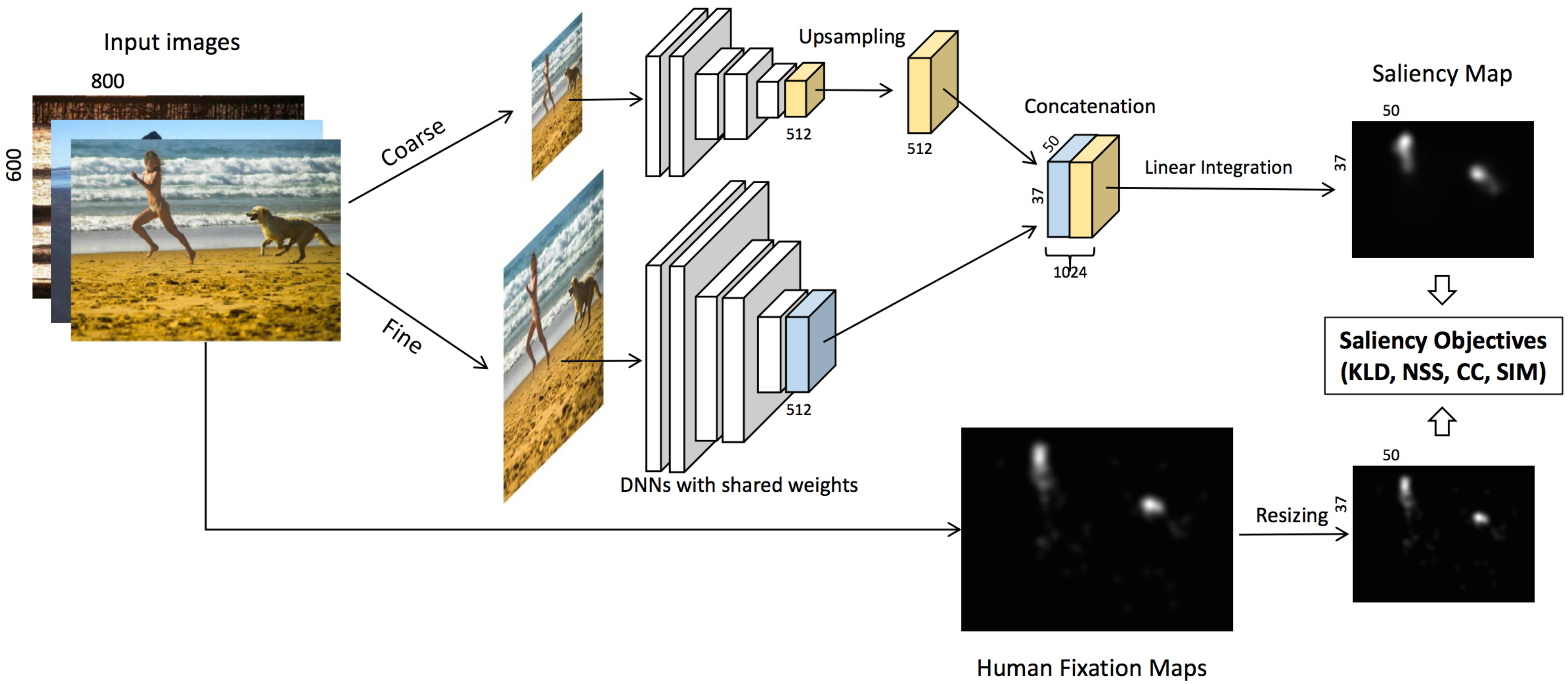
Girshick et al. 14

Applications

- Use a pre-trained CNN as a feature extractor
- **Fine-tune on smaller datasets**
- Train from scratch on big data

Saliency Prediction

Reducing the Semantic Gap in Saliency Prediction by Adapting Neural Networks



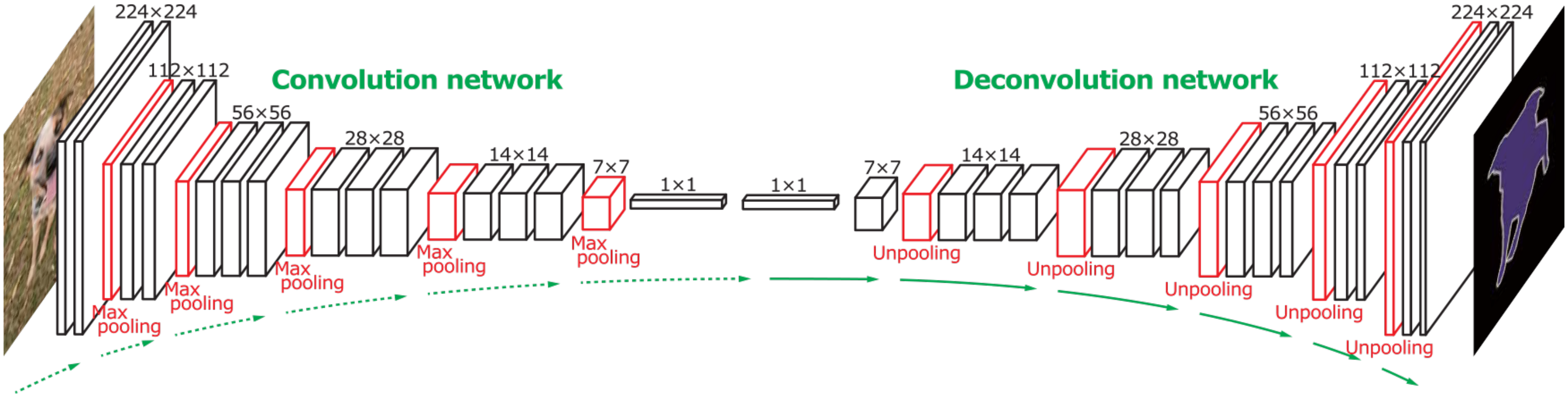
Huang et al. 15

Applications

- Use a pre-trained CNN as a feature extractor
- Fine-tune on limited data
- **Train from scratch on big data**

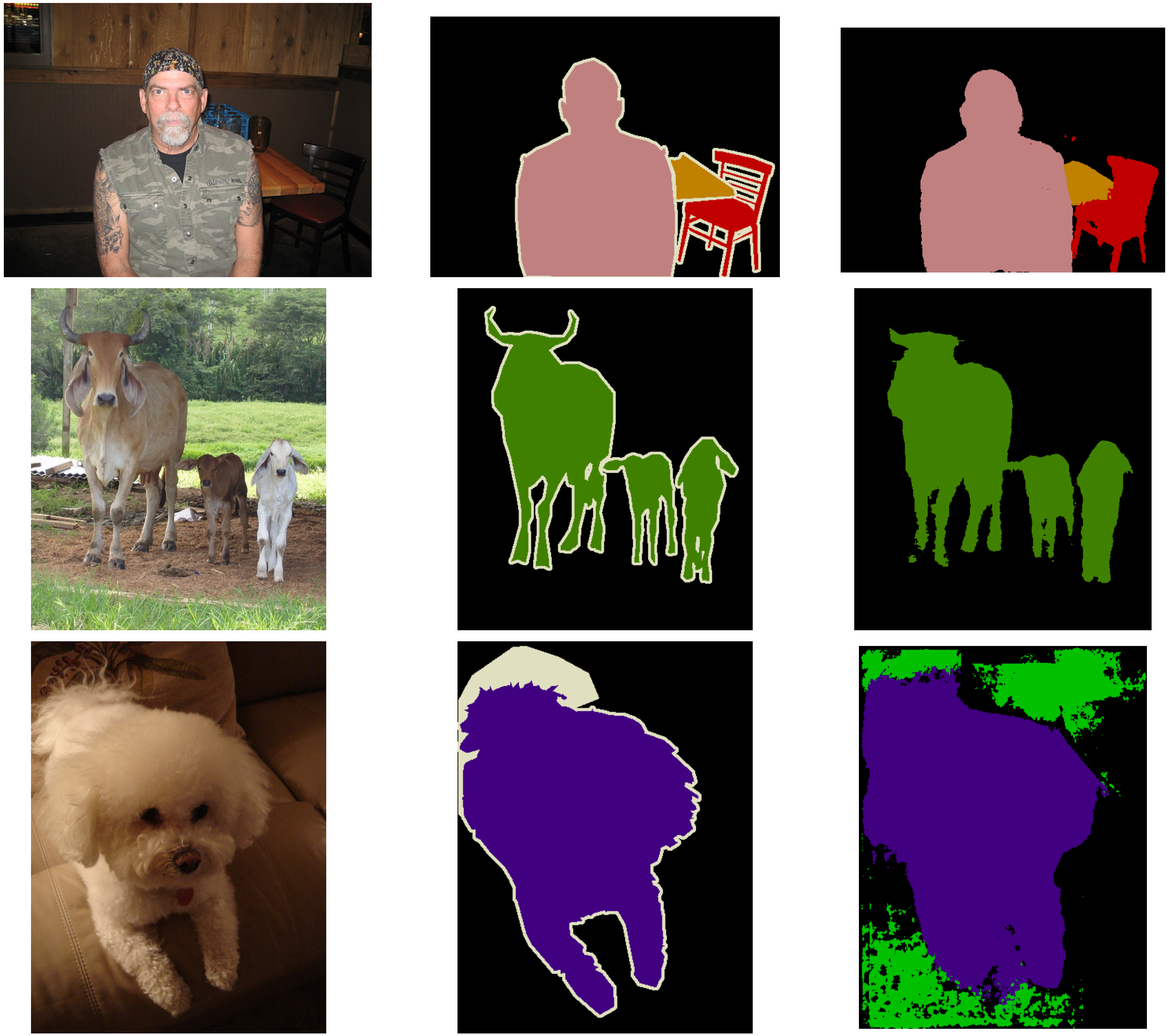
Semantic Segmentation

Learning Deconvolution Network for
Semantic Segmentation



Noh et al. 15

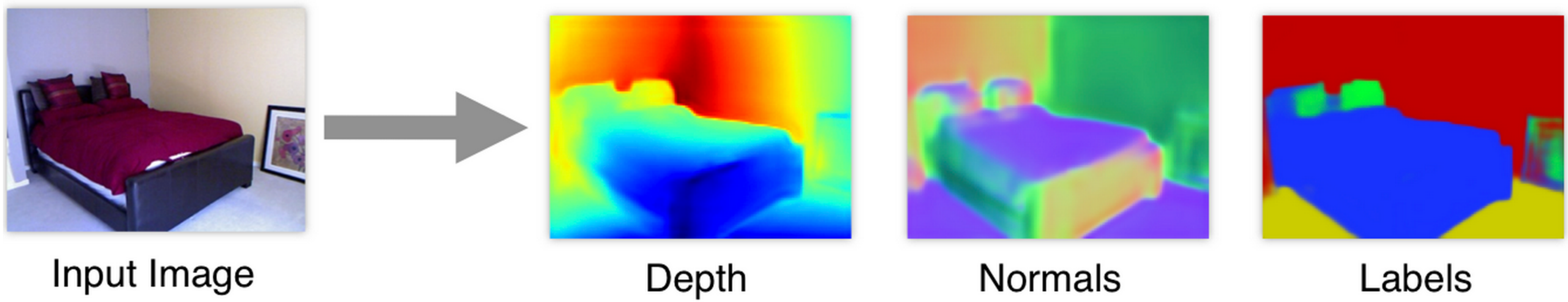
Semantic Segmentation



Noh et al. 15

Depth Map Prediction

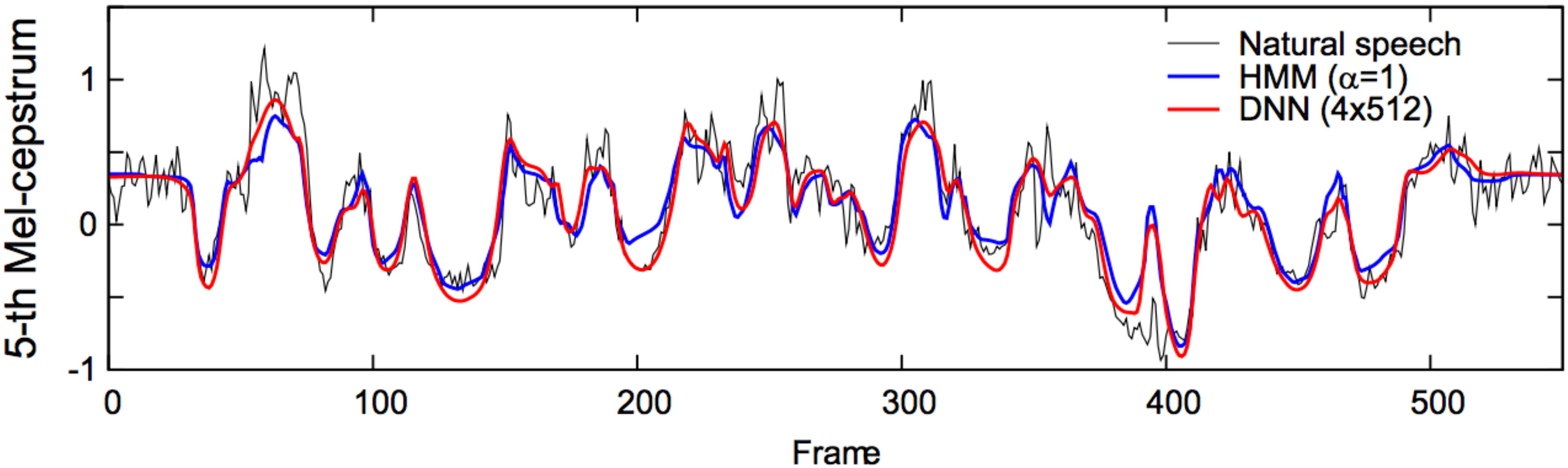
Depth Map Prediction from a Single Image
using a Multi-Scale Deep Network



Eigen and Fergus 14

Applications

not only for vision…

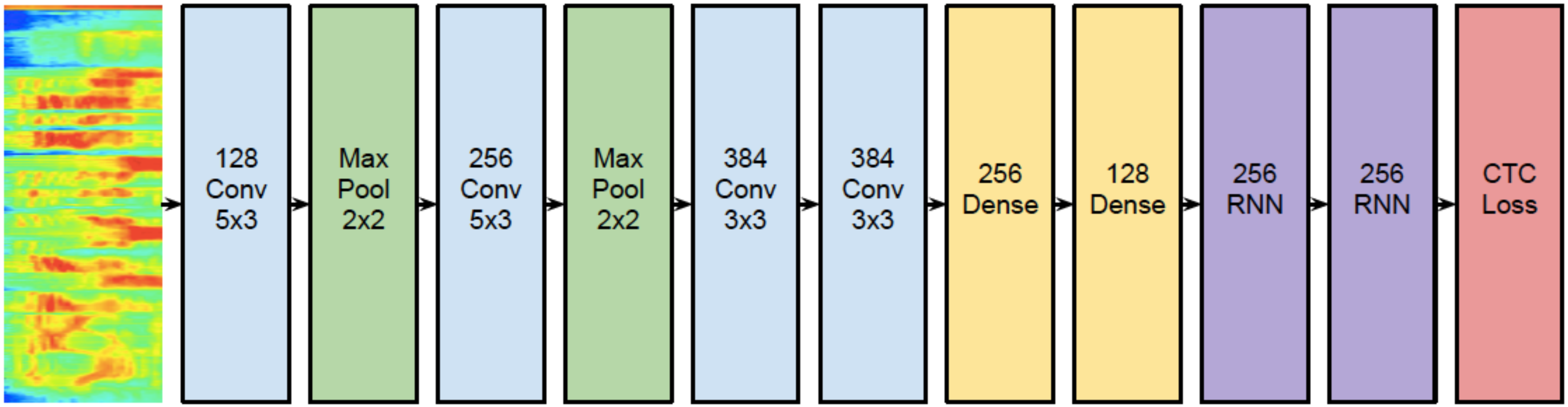


Statistical parametric speech synthesis
using deep neural networks

Zen et. al 13

Applications

End-to-End Deep Neural Network for Automatic Speech Recognition



phonemes recognition

Song and Cai 15

Applications - Frameworks

- Caffe
  - C++ with Matlab and Python interfaces
  - http://caffe.berkeleyvision.org
- Torch
  - Lua
  - http://torch.ch
- Theano
  - Python
  - https://pypi.python.org/pypi/Theano
- MatConvNet
  - Matlab
  - http://www.vlfeat.org/matconvnet/

Overview

- Introduction
- Artificial Neural Networks
- Computational Models of Object Recognition
- Artificial Neural Networks for Object Recognition
- Applications
- Limitations and Open Questions

Open Questions

Adversarial examples

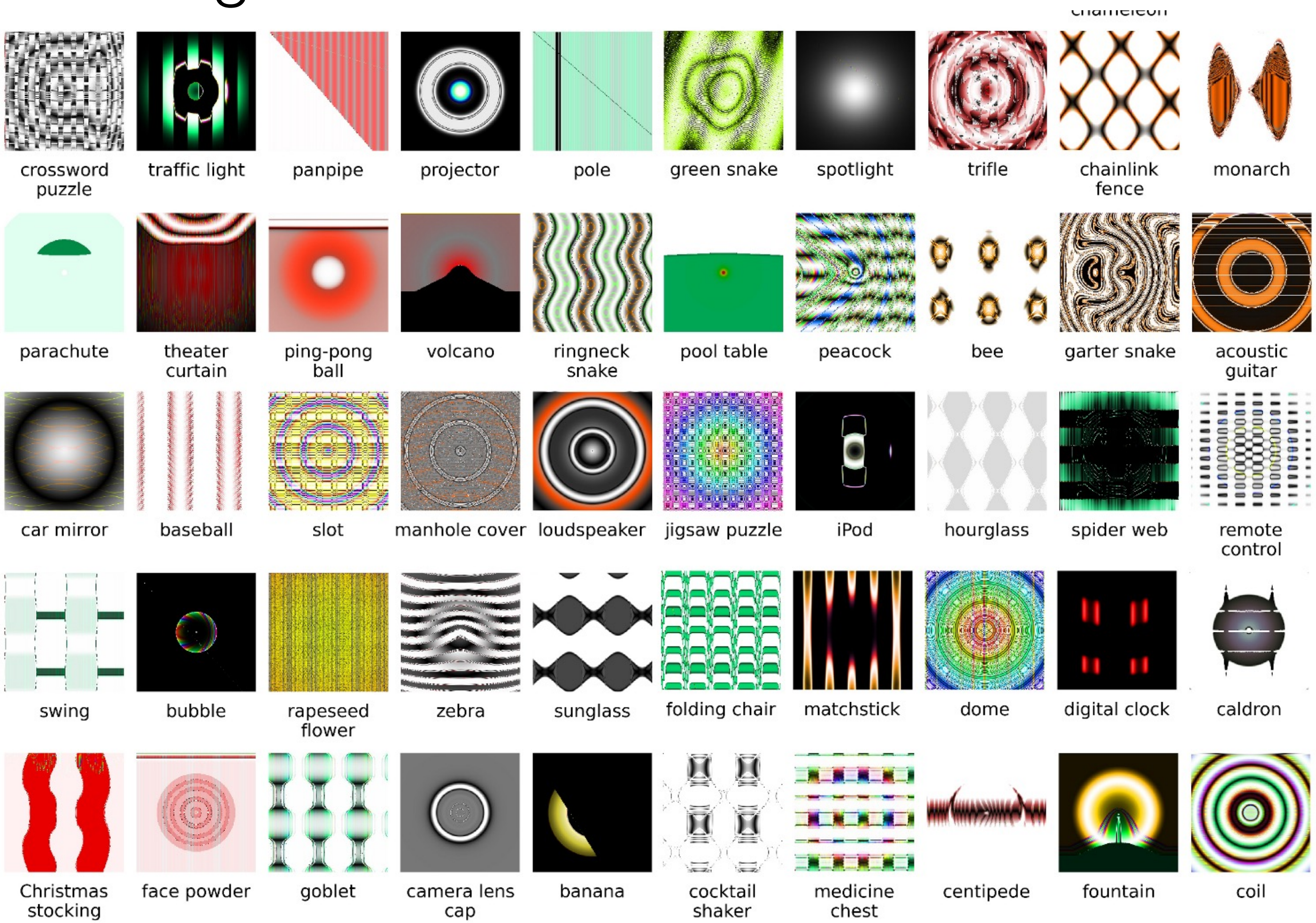
(a) (b)

Figure 5: Adversarial examples generated for AlexNet [9].(Left) is correctly predicted sample, (center) difference between correct image, and image predicted incorrectly magnified by 10x (values shifted by 128 and clamped), (right) adversarial example. Average distortion based on 64 examples is 0.006508.

Szegedy et al. 14

Open Questions

Synthetic images that fool DNN



Nguyen et al. 15

Open Questions

Adversarial examples

Synthetic images that fool DNN

Memory?

Why hierarchies work better than shallow NN?

…

References

Rosenblatt, Frank (1957), The Perceptron--a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. *Learning internal representations by error propagation*. No. ICS-8506. CALIFORNIA UNIV SAN DIEGO LA JOLLA INST FOR COGNITIVE SCIENCE, 1985.

Kobatake, Eucaly, and Keiji Tanaka. "Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex." *Journal of neurophysiology* 71.3 (1994): 856-867.

Hubel, David H., and Torsten N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex." *The Journal of physiology* 148.3 (1959): 574-591.

Riesenhuber, Maximilian, and Tomaso Poggio. "Hierarchical models of object recognition in cortex." *Nature neuroscience* 2.11 (1999): 1019-1025.

Riesenhuber, Maximilian, and Tomaso Poggio. "CBF: A new framework for object categorization in cortex." *Biologically Motivated Computer Vision*. Springer Berlin Heidelberg, 2000.

Serre, Thomas, et al. *A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex*. No. AI MEMO-2005-036. MASSACHUSETTS INST OF TECH CAMBRIDGE MA CENTER FOR BIOLOGICAL AND COMPUTATIONAL LEARNING, 2005.

Serre, Thomas, Aude Oliva, and Tomaso Poggio. "A feedforward architecture accounts for rapid categorization." *Proceedings of the National Academy of Sciences* 104.15 (2007): 6424-6429.

Serre, Thomas, and Maximilian Riesenhuber. *Realistic modeling of simple and complex cell tuning in the HMAX model, and implications for invariant object recognition in cortex*. No. AI-MEMO-2004-017. MASSACHUSETTS INST OF TECH CAMBRIDGE COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LAB, 2004.

LeCun, Yann, et al. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86.11 (1998): 2278-2324.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

References

Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).

Szegedy, Christian, et al. "Going deeper with convolutions." *arXiv preprint arXiv:1409.4842*

Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *Computer Vision–ECCV 2014*. Springer International Publishing, 2014. 818-833.

Simonyan, Karen, Andrea Vedaldi, and Andrew Zisserman. "Deep inside convolutional networks: Visualising image classification models and saliency maps." *arXiv preprint arXiv:1312.6034* (2013).

Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014.

Xun Huang, Chengyao Shen, Xavier Boix, and Qi Zhao. “SALICON: Reducing the Semantic Gap in Saliency Prediction by Adapting Deep Neural Networks”. ICCV 2015

Noh, Hyeonwoo, Seunghoon Hong, and Bohyung Han. "Learning Deconvolution Network for Semantic Segmentation." *arXiv preprint arXiv:*

Eigen, David, Christian Puhrsch, and Rob Fergus. "Depth map prediction from a single image using a multi-scale deep network." *Advances in Neural Information Processing Systems*. 2014.

Ze, Heiga, Alan Senior, and Martin Schuster. "Statistical parametric speech synthesis using deep neural networks." *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013.

Miao, Yajie, Mohammad Gowayyed, and Florian Metze. "EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding." *arXiv preprint arXiv:1507.08240* (2015).

Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." *arXiv preprint arXiv:1412.6572* (2014).

Nguyen, Anh, Jason Yosinski, and Jeff Clune. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images." *arXiv preprint arXiv:1412.1897* (2014).