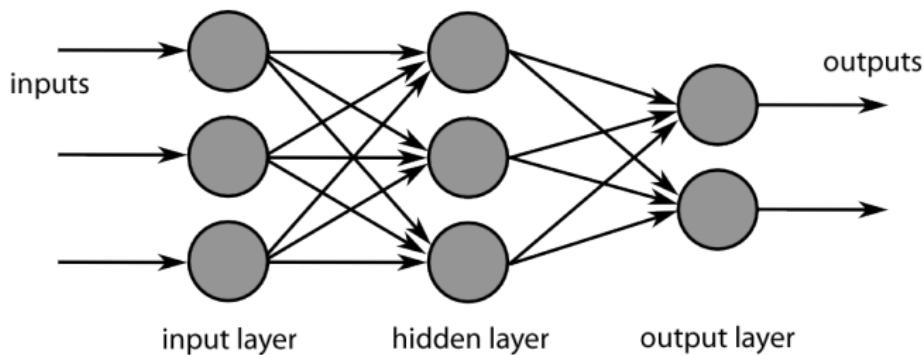


Convolutional Neural Networks.

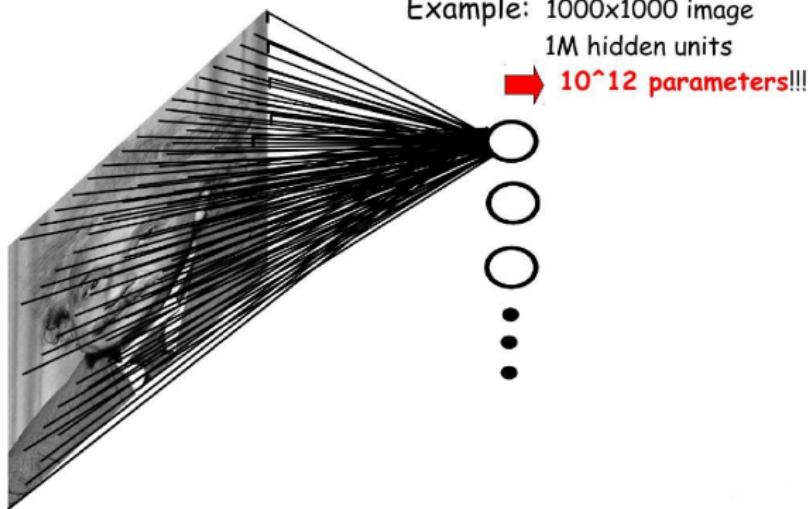
Alvaro Soto

Computer Science Department, PUC

Traditional approach:
Fully connected feed-forward architecture.

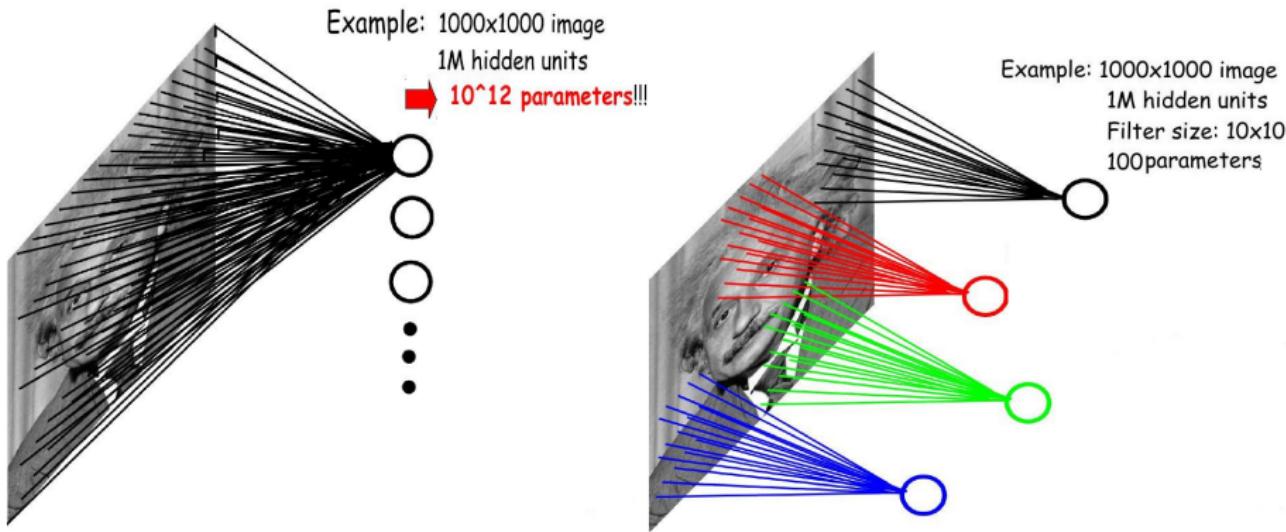


Supervised Neural Nets: Problems



- Huge number of parameters.
- It does not explicitly model local visual pattern (holistic approach).

Solution: Spatially Localized Neurons



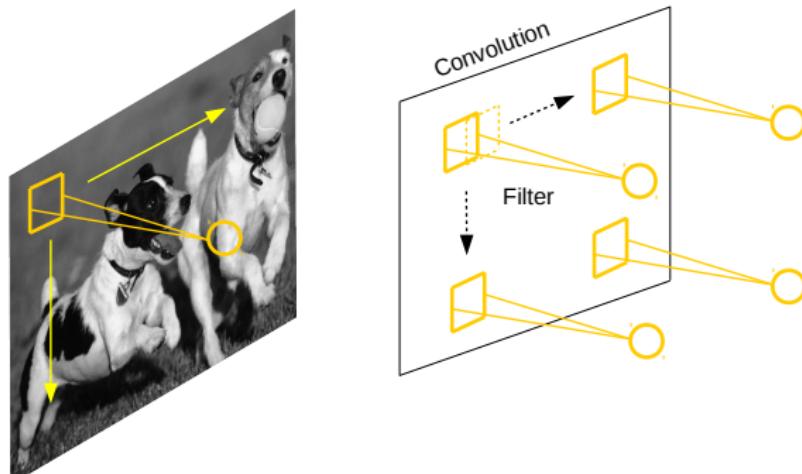
- Spatially localized neurons provide a significant reduction in number of params. In the example, weights associated to each neuron drop from 10^6 to 10^2 .
- Furthermore, these neurons allow the net to capture local image patterns.

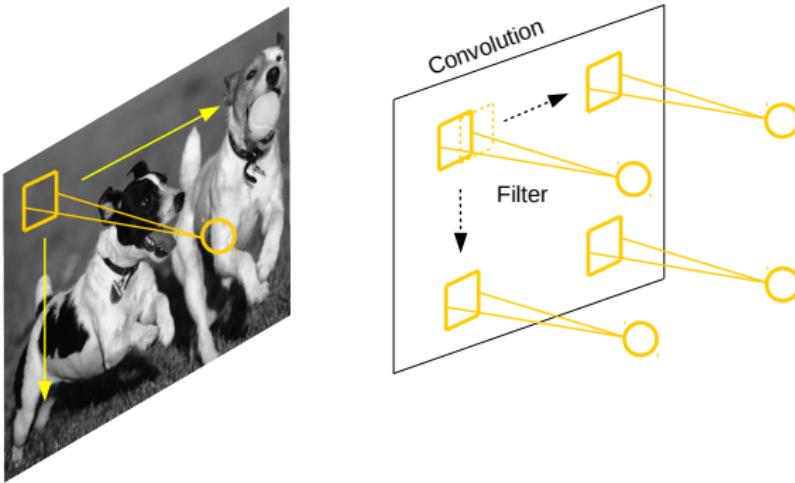
Problem

- A local pattern can potentially occur anywhere in the input.

Solution

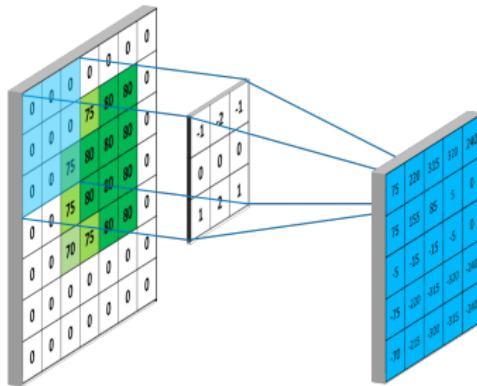
- Convolution provides a mechanism to search for a given pattern (feature) by sliding a set of weights (filter) through the spatial span of the input to generate a "feature map".
- The resulting map consists of the activation of a set of localized neurons that share a set of weights.





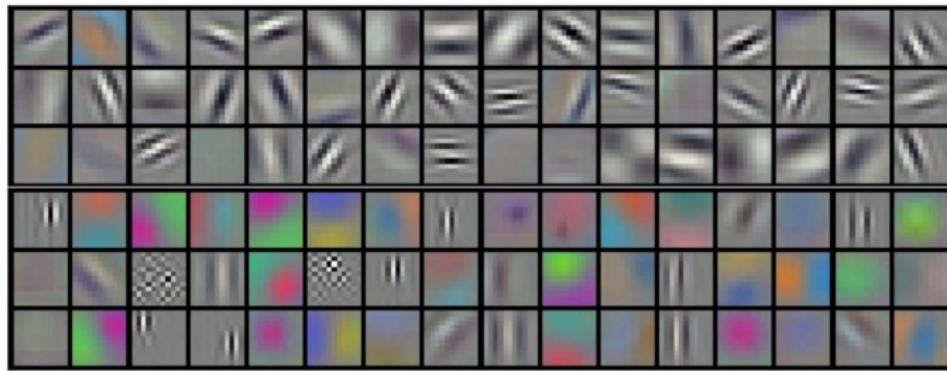
- During training the filter learns to detect a particular pattern (learn?).
- Convolution with each filter generates a particular feature map.
- What type of information contains this feature map?

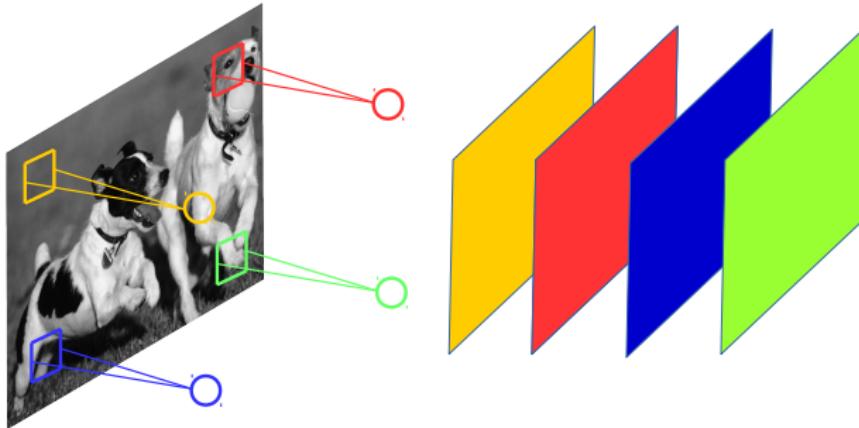
Neuron sharing weights?



- What is the input?
- What is the filter? What is its size?
- What is the feature map?, How many neurons?

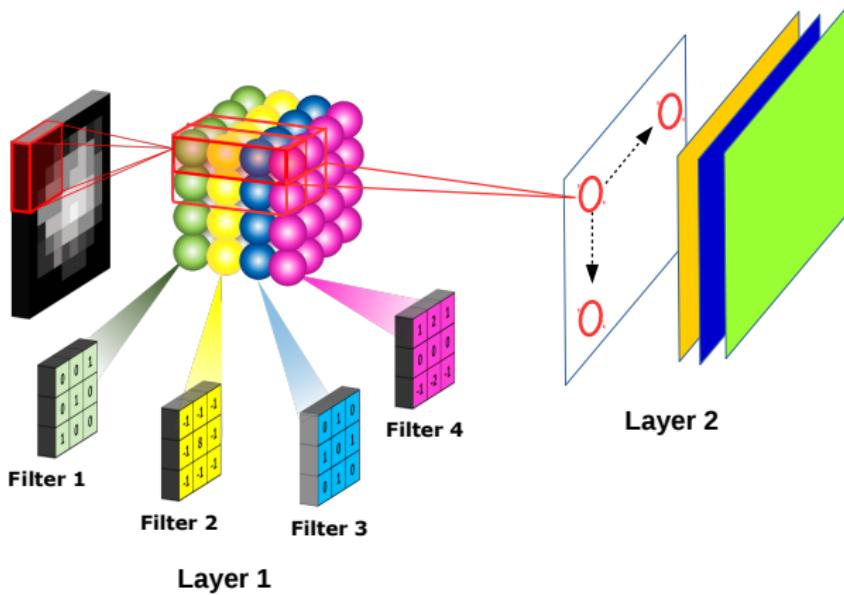
Each filter detects a specific visual pattern



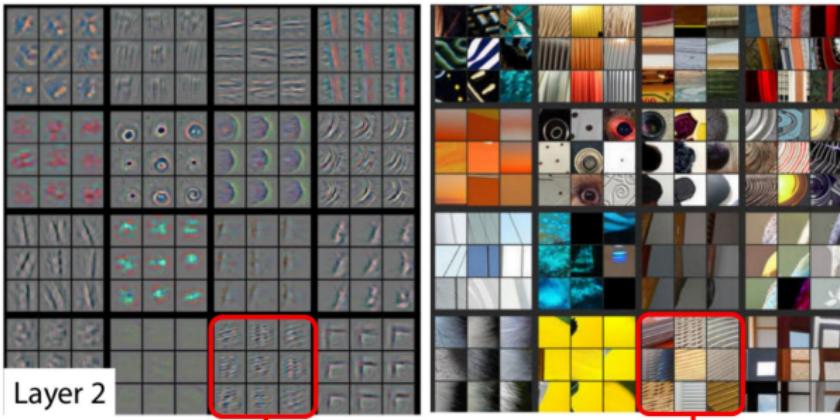


- The convolution process generates an activation map for each filter, which is given by a set of neurons that **share weights**.
- By sharing weights, the model presents an improved mechanism to deal with the spatial variability of local patterns.
- Each set of shared weights allows the net to learn a filter specifically tuned to detect a particular local pattern (filter banks).

A Hierarchy of Network Layers



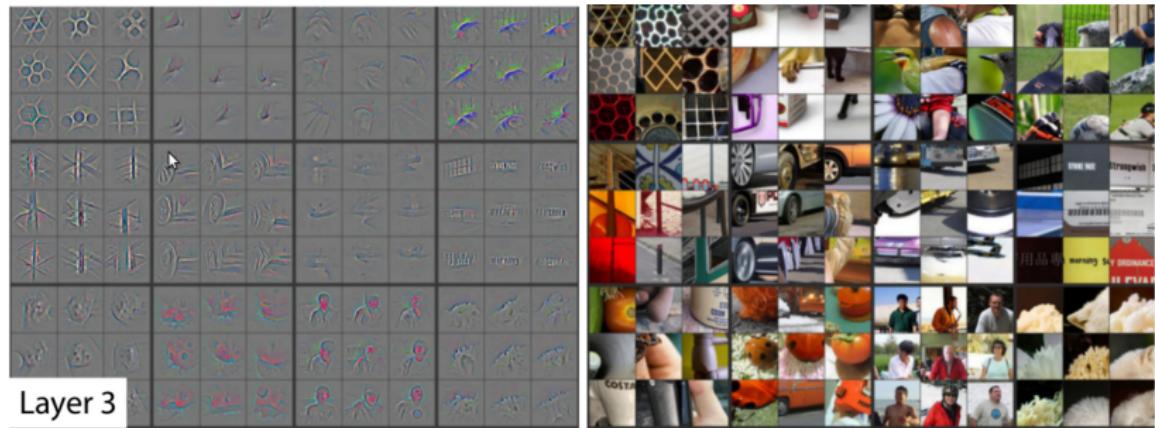
Ex. Visualization of Neuron Activations



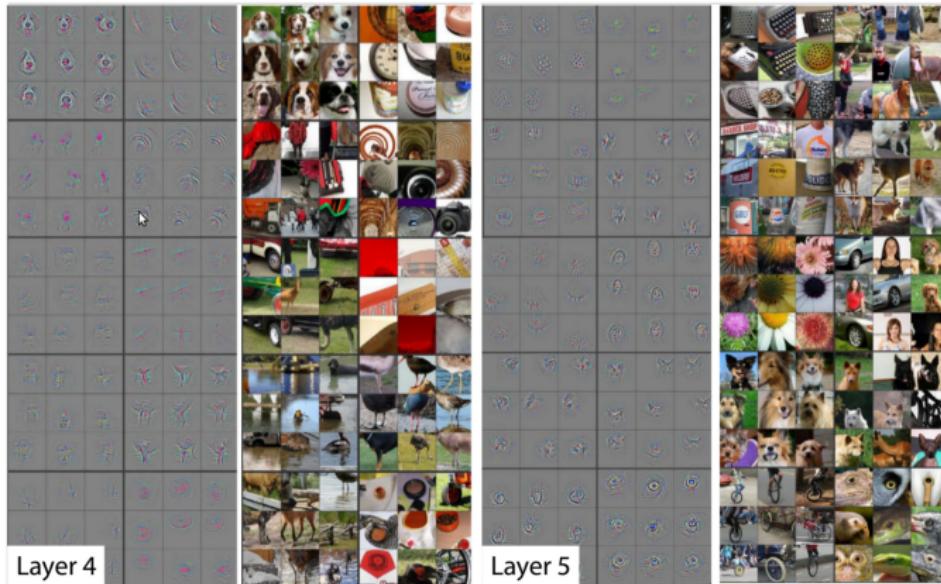
Reconstruction of image patches using activation of a layer 2 neuron (indicates aspect of patches which unit is sensitive to).

Top 9 image patches that cause maximal activation of a layer 2 neuron.

Ex. Visualization of Neuron Activations

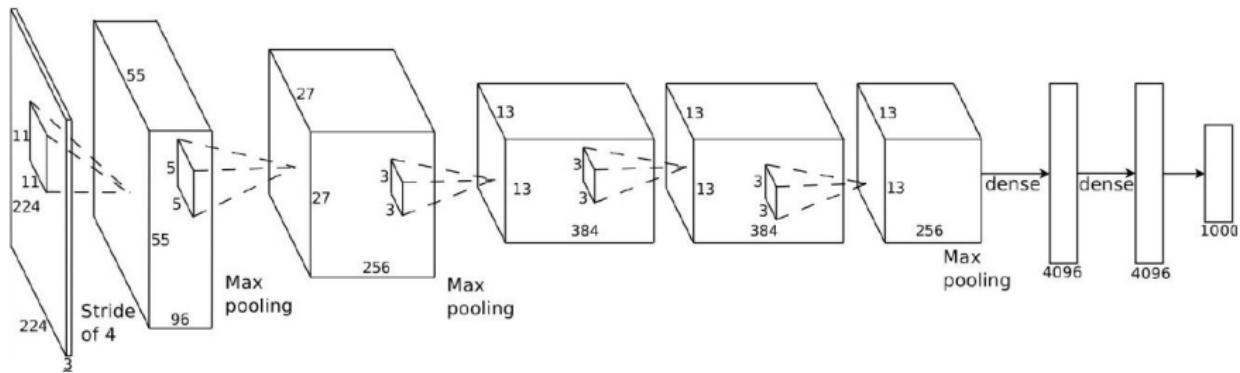


Ex. Visualization of Neuron Activations



Going Deeper: Stack Several Layers of CNNs

Example: Meet Alex's Net



- The role of each convolutional layer is to detect local conjunctions of features from the previous layer.

Case Example

ImageNet Classification with Deep Convolutional Neural Networks

A. Krizhevsky, I. Sutskever, and G. E. Hinton (NIPS-2012)

- 15M images.
- 22K categories.
- Images collected from Web.
- Human labelers (Amazon's Mechanical Turk crowd-sourcing).



ImageNet Large Scale Visual Recognition Challenge (ILSVRC-2010)

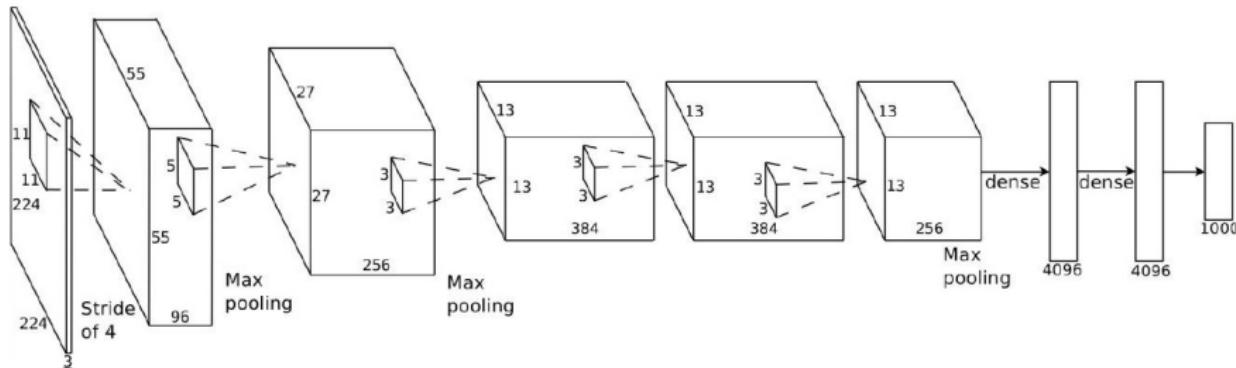
ILSVRC-2010

- 1K categories.
- 1.2M training images (1000 per category).
- 50K validation images.
- 150K testing images.
- RGB images.
- Variable-resolution, but in this application images are scaled to 256x256 pixels.

Classification goals

- Make 1 guess about the label (Top-1 error).
- make 5 guesses about the label (Top-5 error).

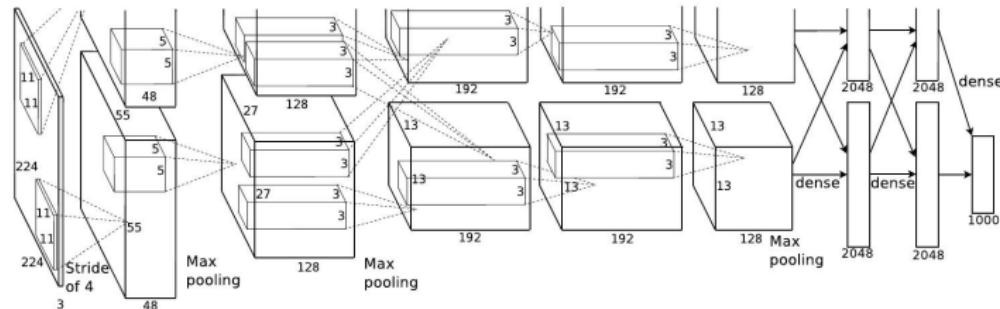
AlexNet Arquitecture:



Example Layer-1:

- nFilters = 96, Size = 11×11 , Depth = 3.
- nParams (weights) = $96 \times 11 \times 11 \times 3 = 34848$.
- Stride = 4, max-pooling = 3x3:1, nNeurons (outputs) = $55 \times 55 \times 96 = 290400$.
(obs: they use 3 zero-padding, then $(227 - 11)/4 + 1 = 55$)

Arquitecture: Trained on 2 GPUs



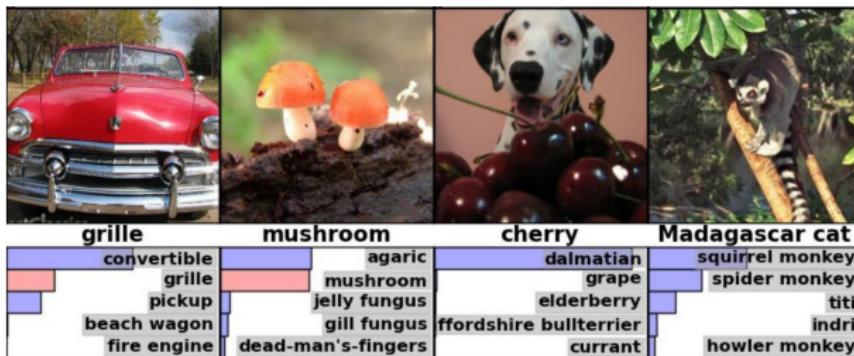
- Trained with stochastic gradient descent on two NVIDIA GTX 580 3GB GPUs for about a week.
- Approx. size:
 - Aprox. 650,000 neurons.
 - 60,000,000 parameters.
 - 630,000,000 connections.

Results on ILSVRC

Results on ILSVRC-2012 competition:

- Top-5 error rate: 15.3%.
- 2nd best team: 26.2% error.
- Human error approx. 8%.

Take these numbers with care, as an example, the following cases correspond to wrong classifications:





lens cap

reflex camera
Polaroid camera
pencil sharpener
switch
combination lock



abacus

typewriter keyboard
space bar
computer keyboard
accordion



slug

zucchini
ground beetle
common newt
water snake



hen

cock
cocker spaniel
partridge
English setter



tiger

tiger
tiger cat
tabby
boxer
Saint Bernard



chambered nautilus

lampshade
throne
goblet
table lamp
hamper



tape player

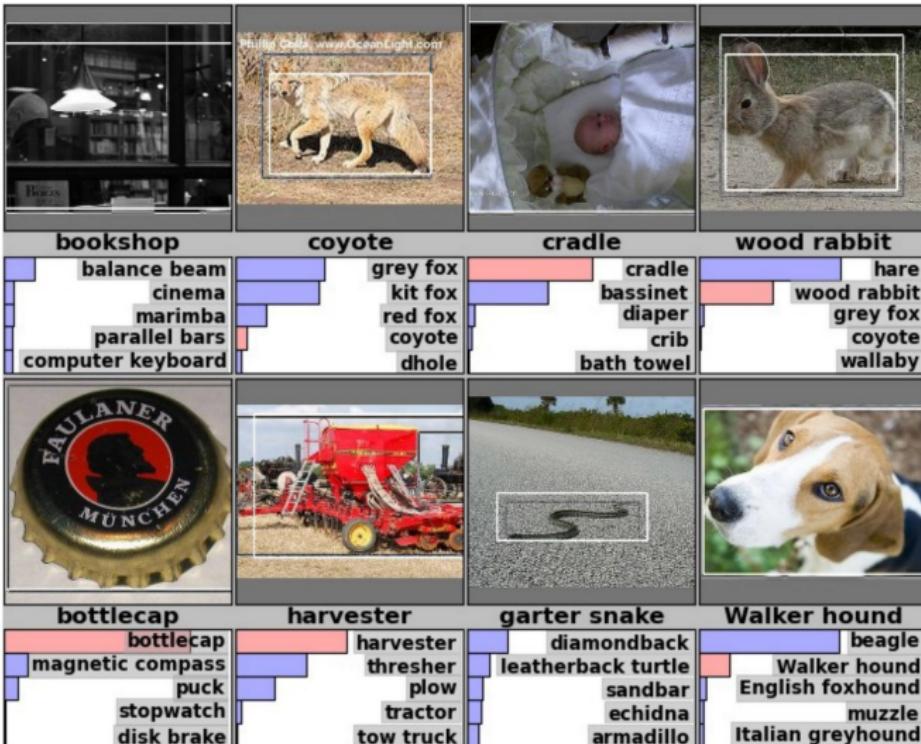
cellular telephone
slot
reflex camera
dial telephone
iPod



planetarium

dome
mosque
radio telescope
steel arch bridge





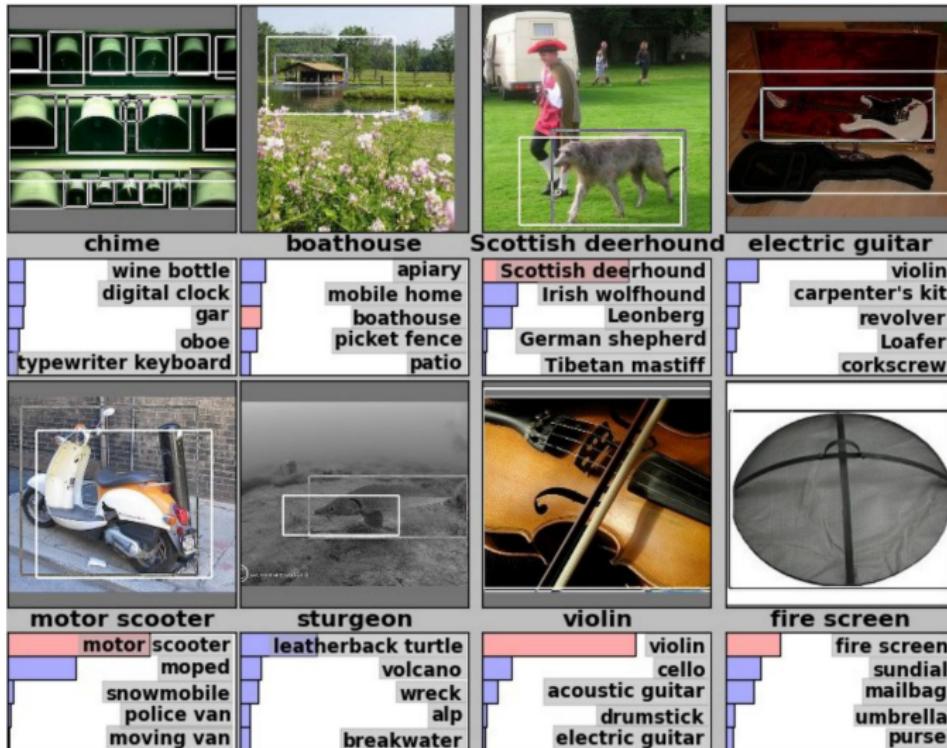


Image Retrieval

First column contains query images from ILSVRC-2010 test set, remaining columns contain retrieved images from training set.

