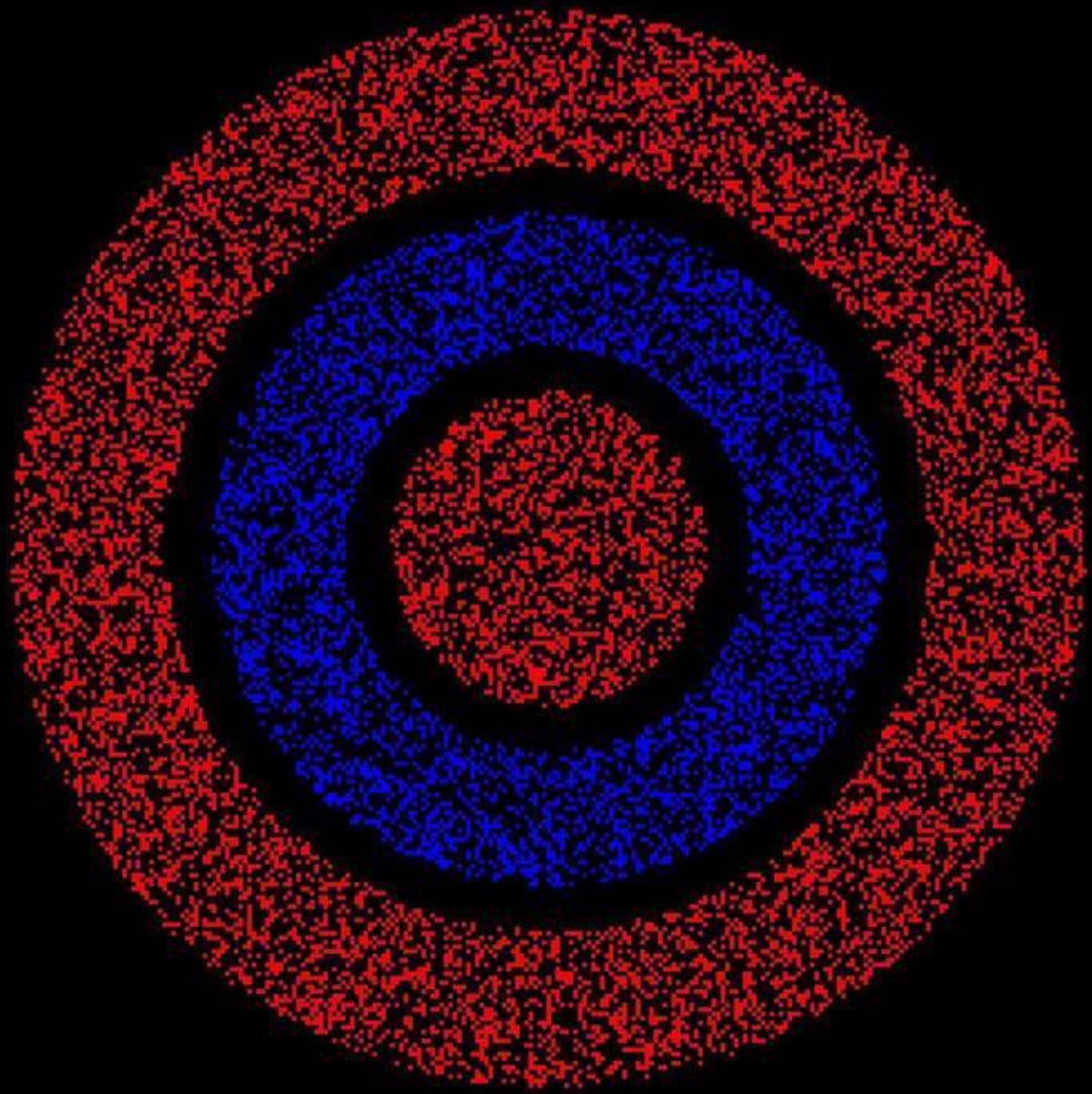
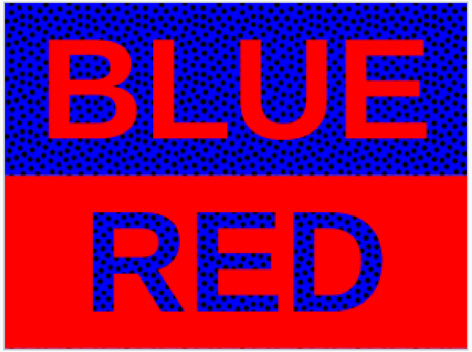


Chromostereopsis

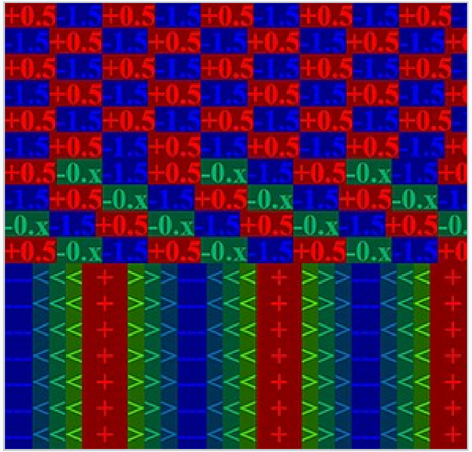
From Wikipedia, the free encyclopedia



Blue–red contrast demonstrating depth perception effects

Chromostereopsis is a visual illusion whereby the impression of depth is conveyed in two-dimensional color images, usually of red–blue or red–green colors, but can also be perceived with red–grey or blue–grey images.[1][2] Such illusions have been reported for over a century and have generally been attributed to some form of chromatic aberration.[3][4][5][6][7]

Chromatic aberration results from the differential refraction of light depending on its wavelength, causing some light rays to converge before others in the eye (longitudinal chromatic aberration or LCA) and/or to be located on non-corresponding locations of the two eyes during binocular viewing (transverse chromatic aberration or TCA).



3 Layers of depths "Rivers, Valleys & Mountains"

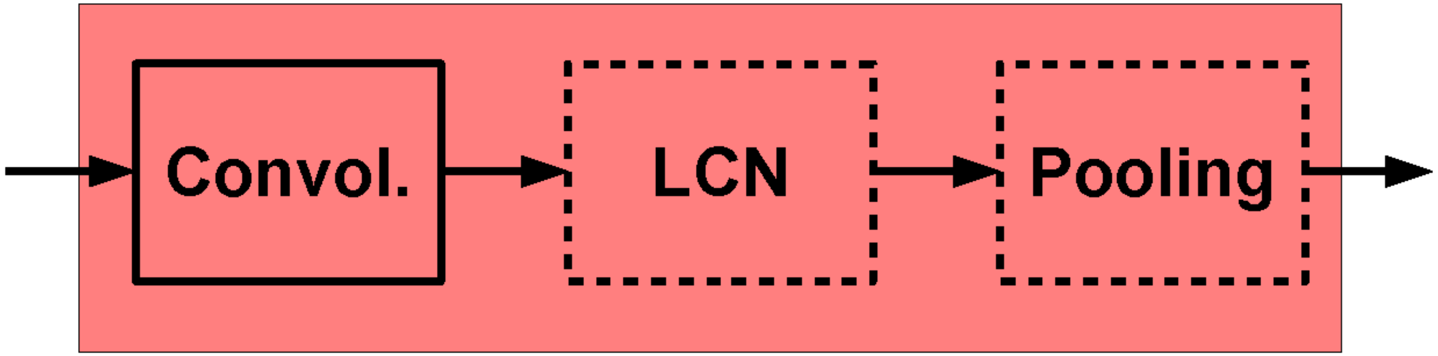
Chromostereopsis is usually observed using a target with red and blue bars and an achromatic background. Positive chromostereopsis is exhibited when the red bars are perceived in front of the blue and negative chromostereopsis is exhibited when the red bars are perceived behind the blue.[8] Several models have been proposed to explain this effect which is often attributed to longitudinal and/or transverse chromatic aberrations.[6] However, some work attributes most of the stereoptic effect to transverse chromatic aberrations in combination with cortical factors.[1][5][7]

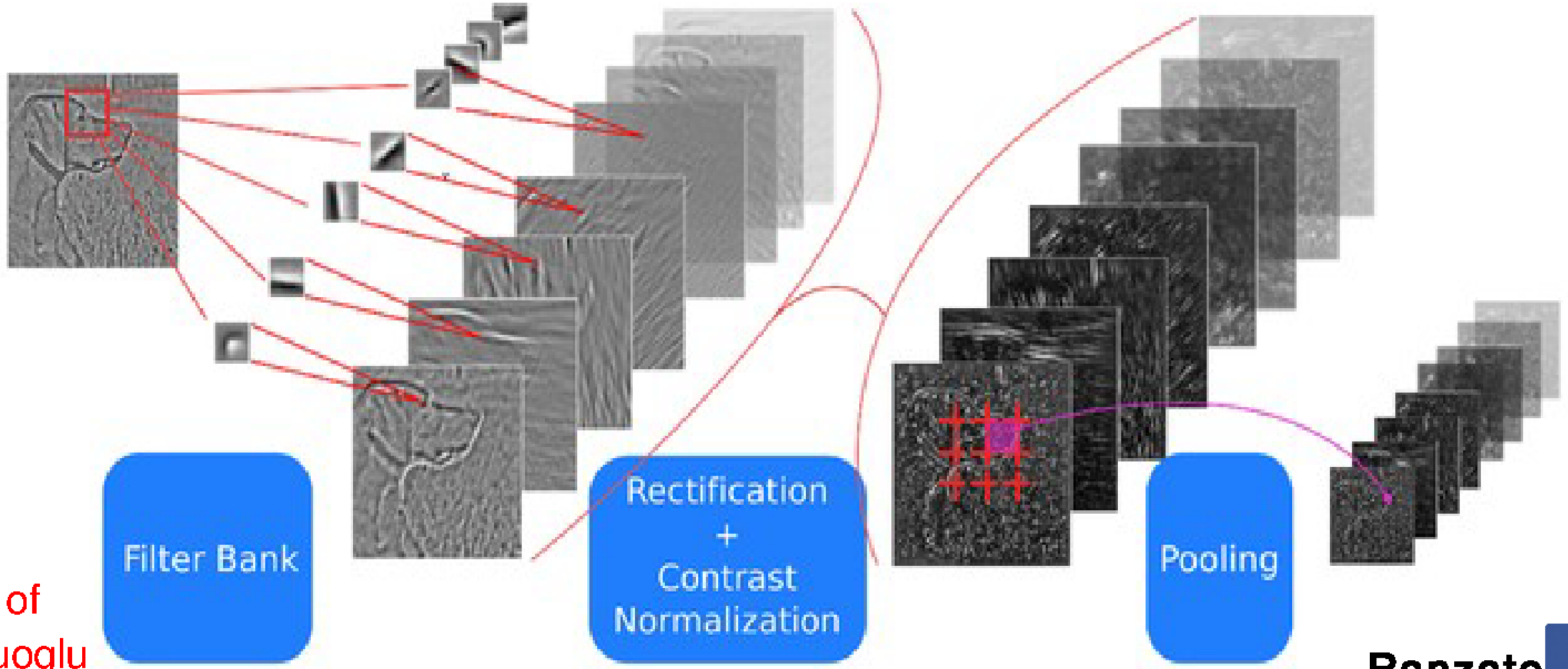
It has been proposed that chromostereopsis could have evolutionary implications in the development of eyespots in certain butterfly species.

The perceived differences in color's optical power span about 2 diopter (Blue: −1.5, Red

ConvNets: Typical Stage

One stage (zoom)



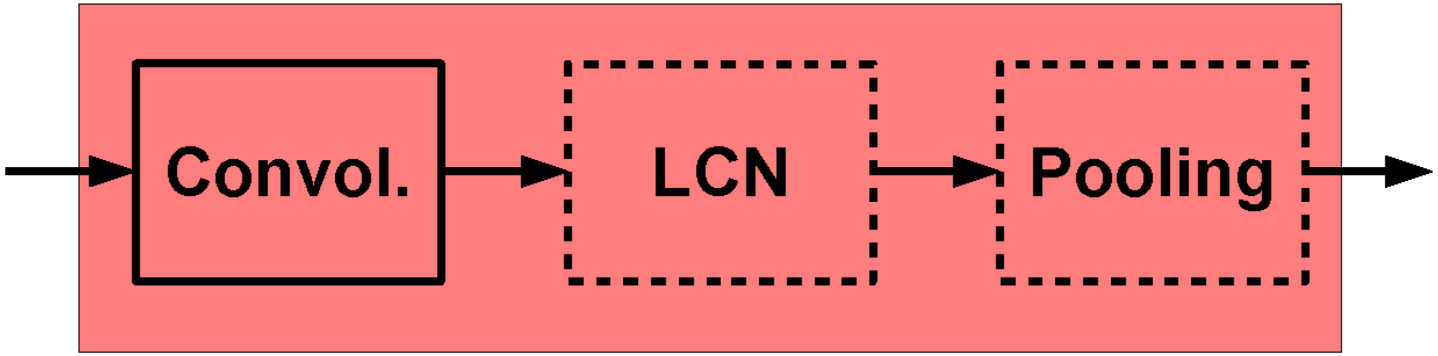


courtesy of
K. Kavukcuoglu

Ranzato

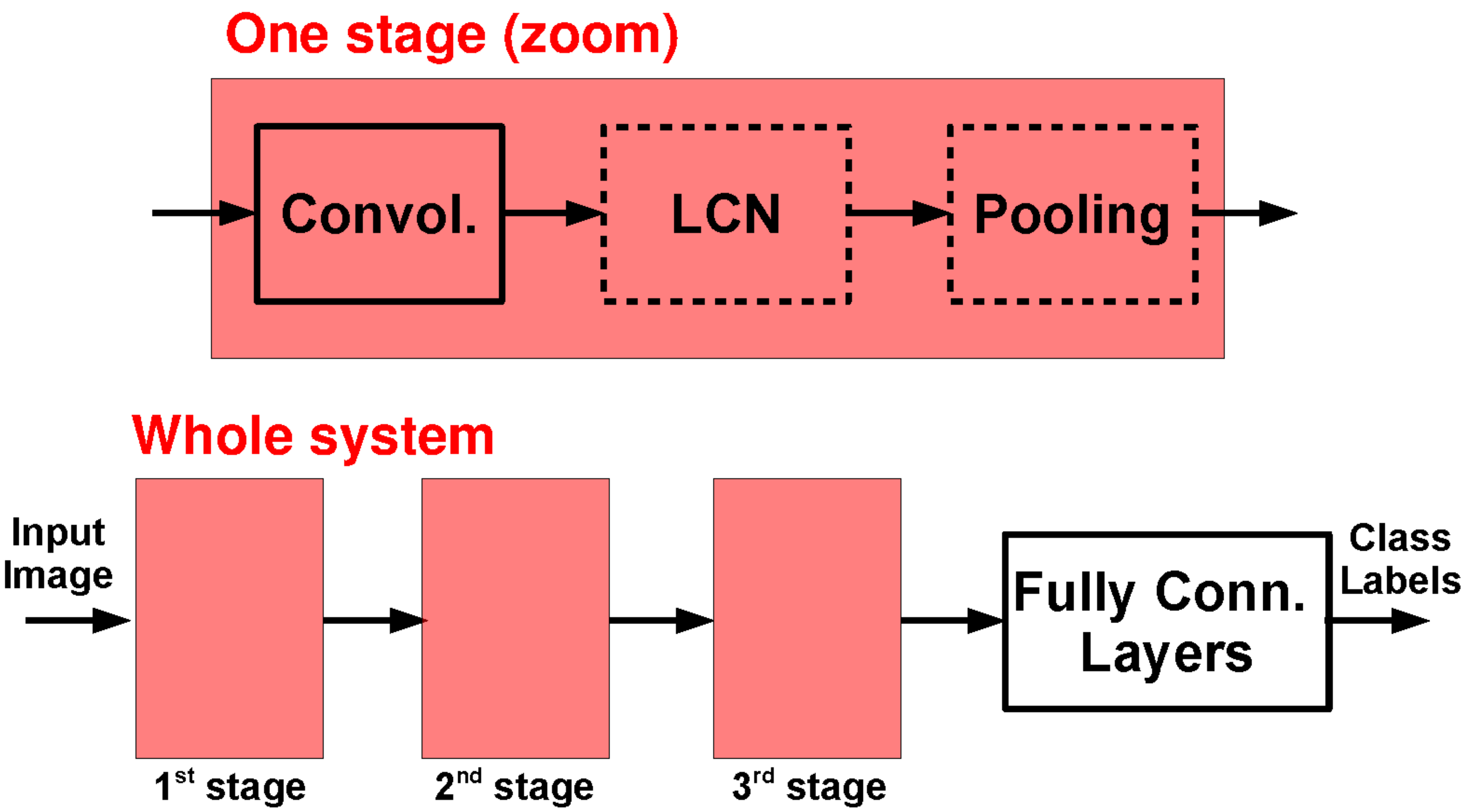
ConvNets: Typical Stage

One stage (zoom)

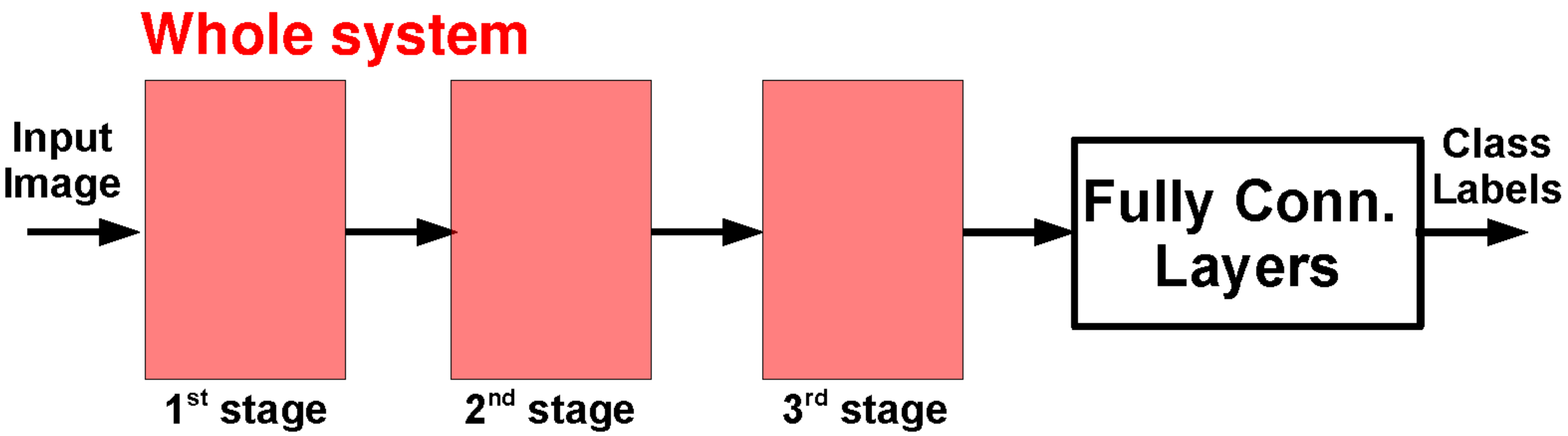


Conceptually similar to: SIFT, HoG, etc.

ConvNets: Typical Architecture



ConvNets: Typical Architecture



Conceptually similar to:

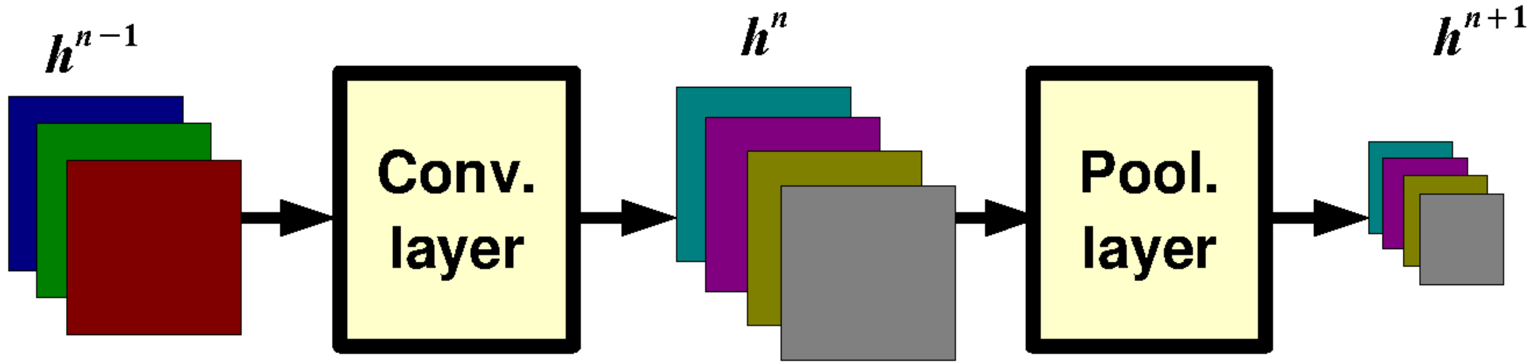
SIFT → K-Means → Pyramid Pooling → SVM

Lazebnik et al. "...Spatial Pyramid Matching..." CVPR 2006

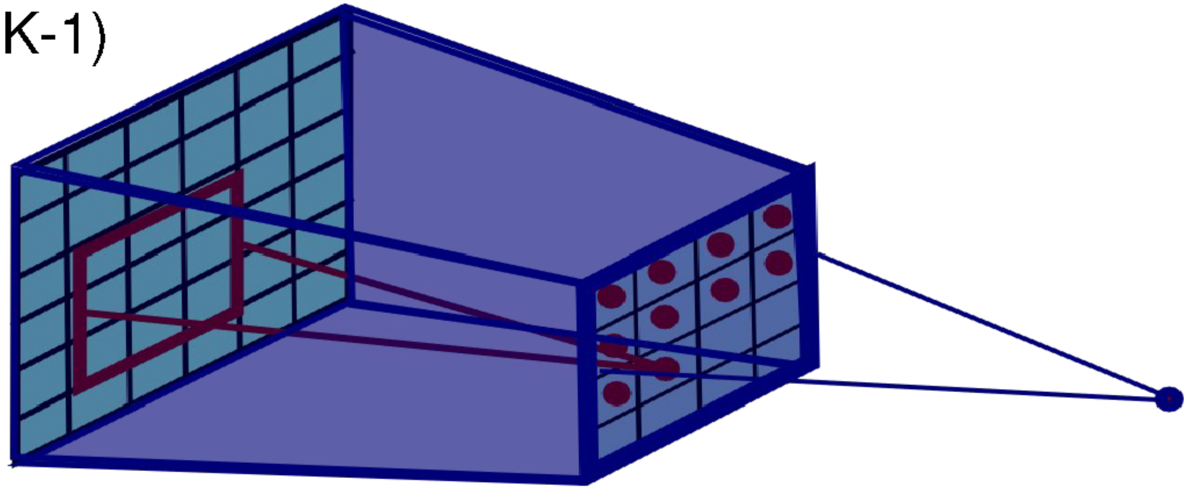
SIFT → Fisher Vect. → Pooling → SVM

Sanchez et al. "Image classifcation with F.V.: Theory and practice" IJCV 2012

Pooling Layer: Receptive Field Size



If convolutional filters have size KxK and stride 1, and pooling layer has pools of size PxP, then each unit in the pooling layer depends upon a patch (at the input of the preceding conv. layer) of size: (P+K-1)x(P+K-1)



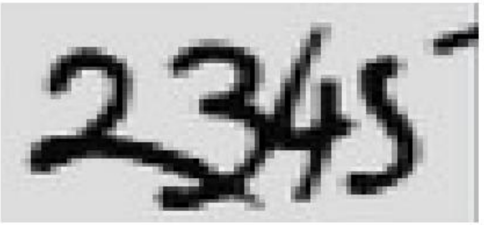
Ranzato

Outline

- Supervised Neural Networks
- Convolutional Neural Networks
- Examples
- Tips

CONV NETS: EXAMPLES

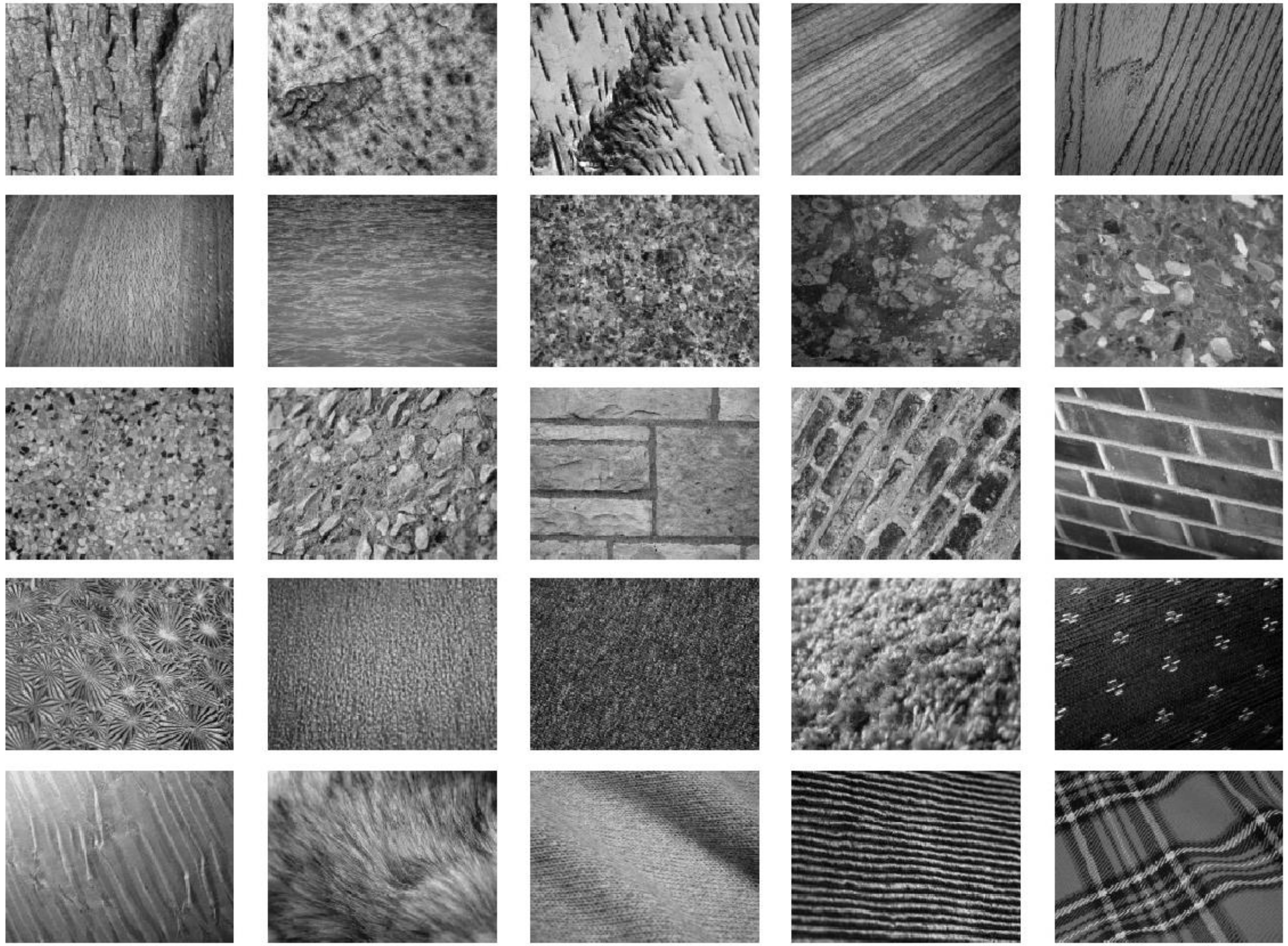
- **OCR / House number & Traffic sign classification**



Ciresan et al. “MCDNN for image classification” CVPR 2012
Wan et al. “Regularization of neural networks using dropconnect” ICML 2013
Jaderberg et al. “Synthetic data and ANN for natural scene text recognition” arXiv 2014

CONV NETS: EXAMPLES

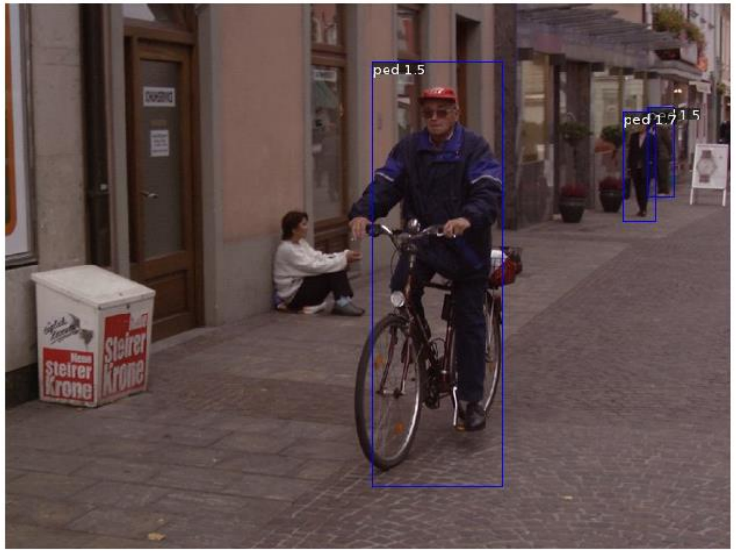
- Texture classification

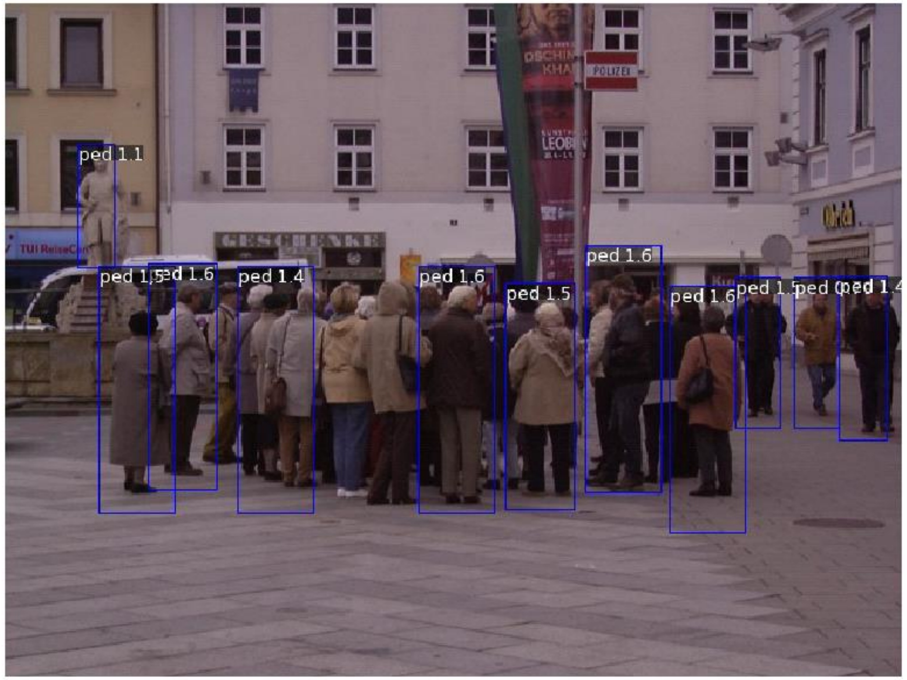


Sifre et al. “Rotation, scaling and deformation invariant scattering...” CVPR 2013

CONV NETS: EXAMPLES

- **Pedestrian detection**

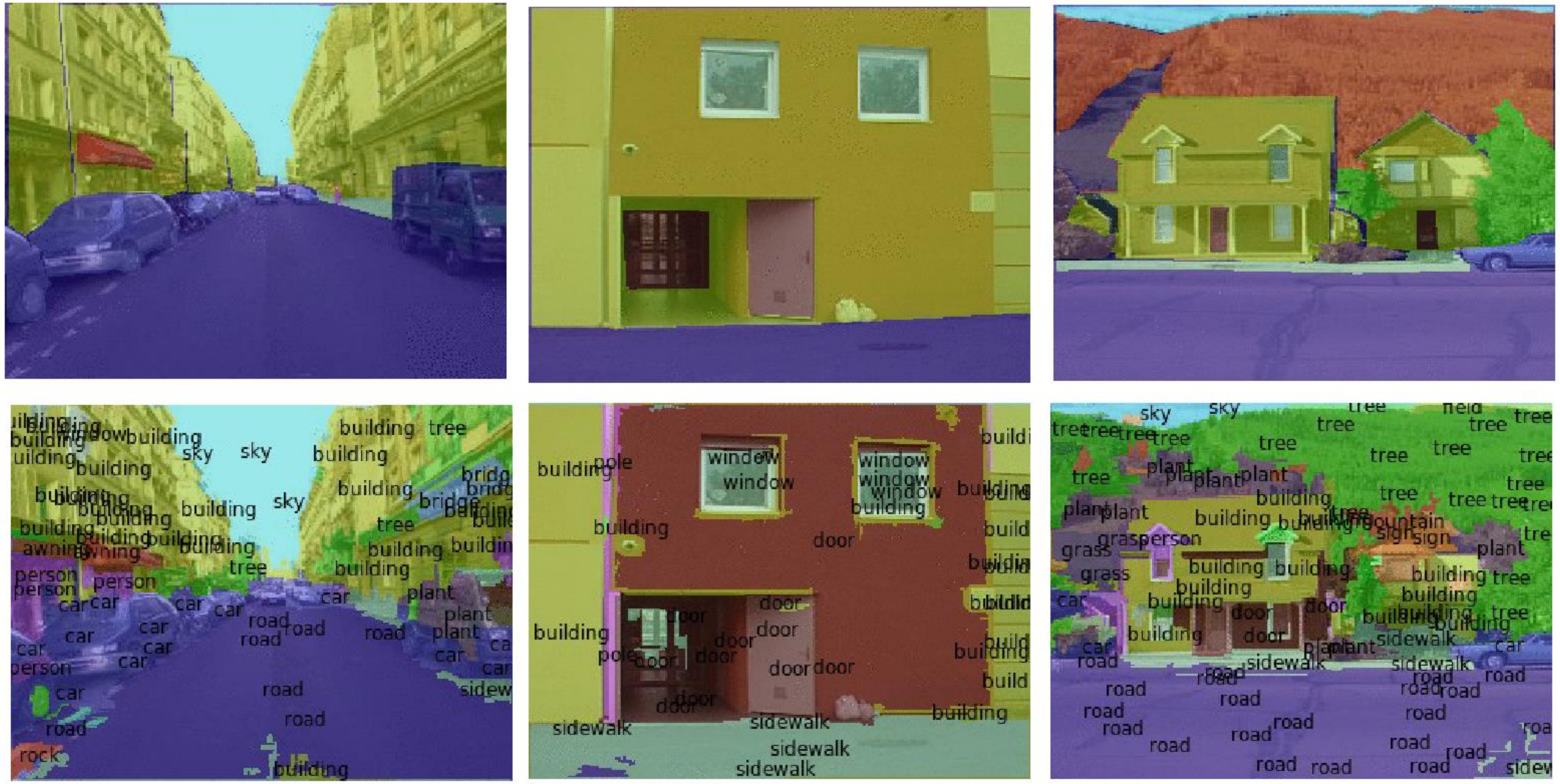




Sermanet et al. "Pedestrian detection with unsupervised multi-stage.." CVPR 2013

CONV NETS: EXAMPLES

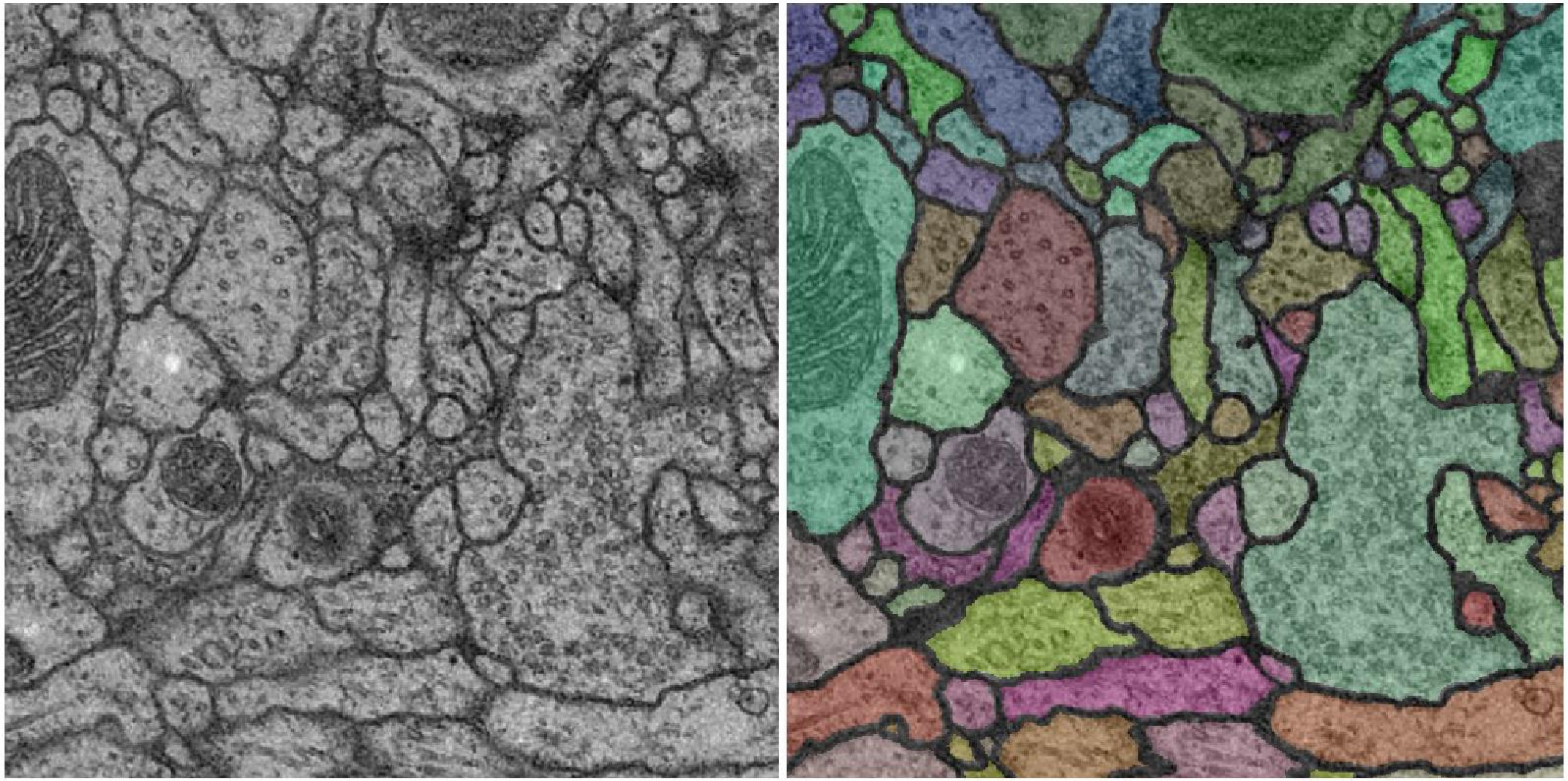
- **Scene Parsing**



Farabet et al. “Learning hierarchical features for scene labeling” PAMI 2013
Pinheiro et al. “Recurrent CNN for scene parsing” arxiv 2013

CONV NETS: EXAMPLES

- **Segmentation 3D volumetric images**



Ciresan et al. "DNN segment neuronal membranes..." NIPS 2012
Turaga et al. "Maximin learning of image segmentation" NIPS 2009

CONV NETS: EXAMPLES

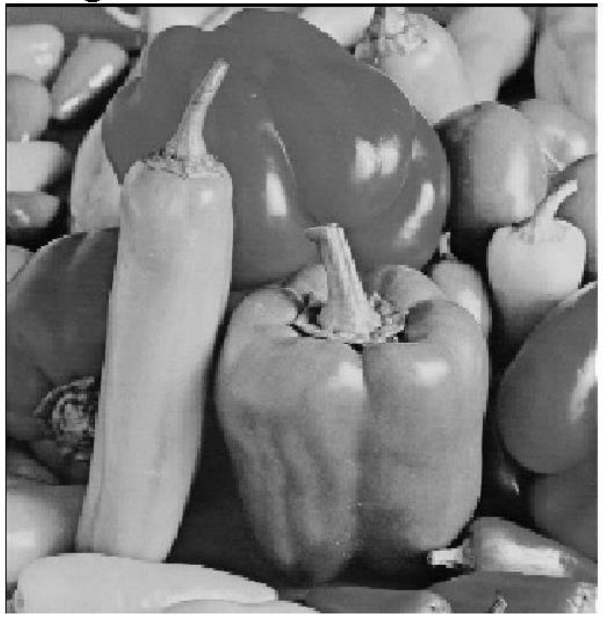
- **Action recognition from videos**

Taylor et al. “Convolutional learning of spatio-temporal features” ECCV 2010
Karpathy et al. “Large-scale video classification with CNNs” CVPR 2014
Simonyan et al. “Two-stream CNNs for action recognition in videos” arXiv 2014

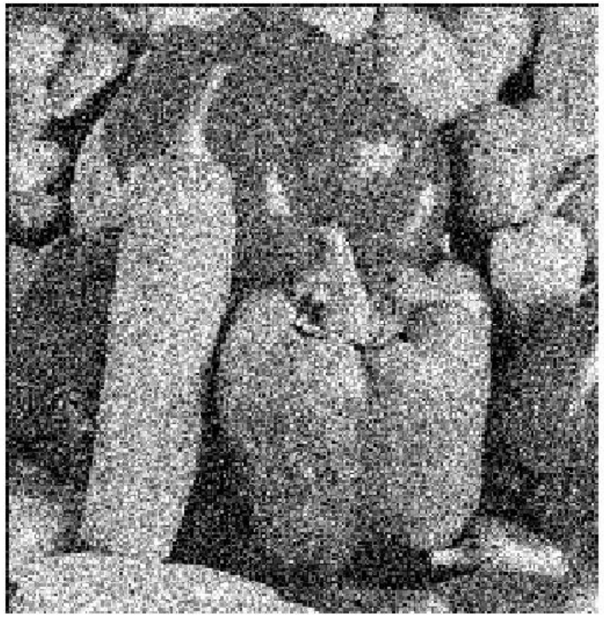
CONV NETS: EXAMPLES

- **Denoising**

original



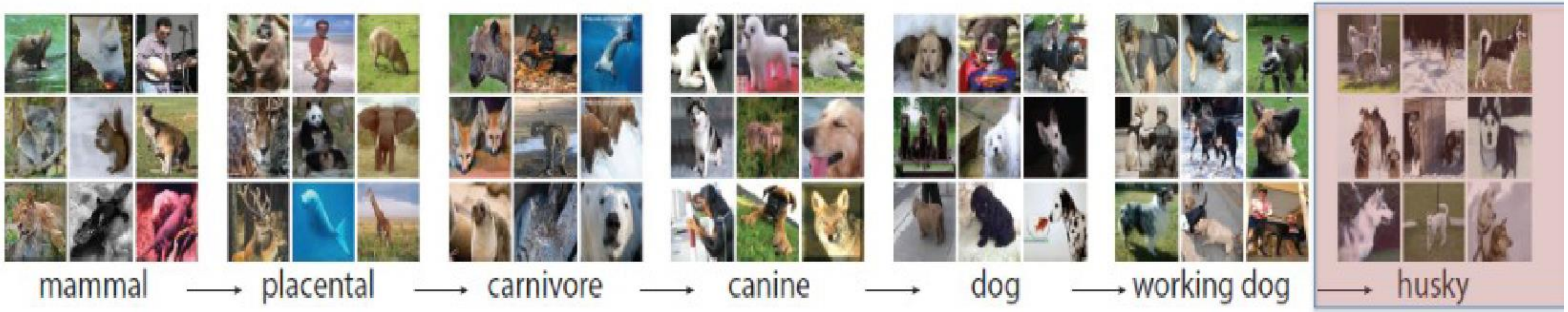
noised



denoised

Burger et al. “Can plain NNs compete with BM3D?” CVPR 2012

Dataset: ImageNet 2012



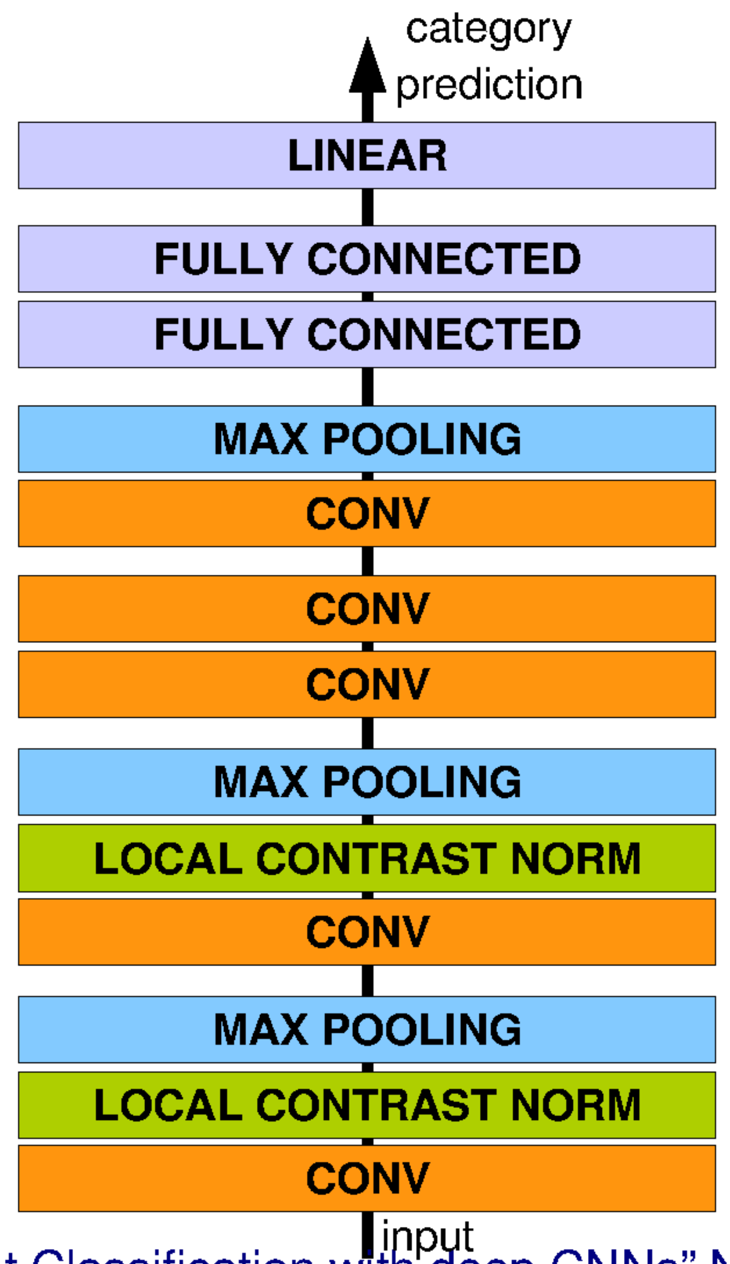
- S: (n) Eskimo dog, **husky** (breed of heavy-coated Arctic sled dog)
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) working dog (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
      - S: (n) dog, domestic dog, Canis familiaris (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) *"the dog barked all night"*
        - S: (n) canine, canid (any of various fissiped mammals with nonretractile claws and typically long muzzles)
          - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have four or five clawed digits on each limb"*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a placenta; all mammals except monotremes and marsupials)
              - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
                - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
                  - S: (n) chordate (any animal of the phylum Chordata having a notochord or spinal column)
                    - S: (n) animal, animate being, beast, brute, creature, fauna (a living organism characterized by voluntary movement)
                      - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
                        - S: (n) living thing, animate thing (a living (or once living) entity)
                          - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                            - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                              - S: (n) physical entity (an entity that has physical existence)
                                - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Deng et al. “Imagenet: a large scale hierarchical image database” CVPR 2009

ImageNet

Examples of hammer:

Architecture for Classification

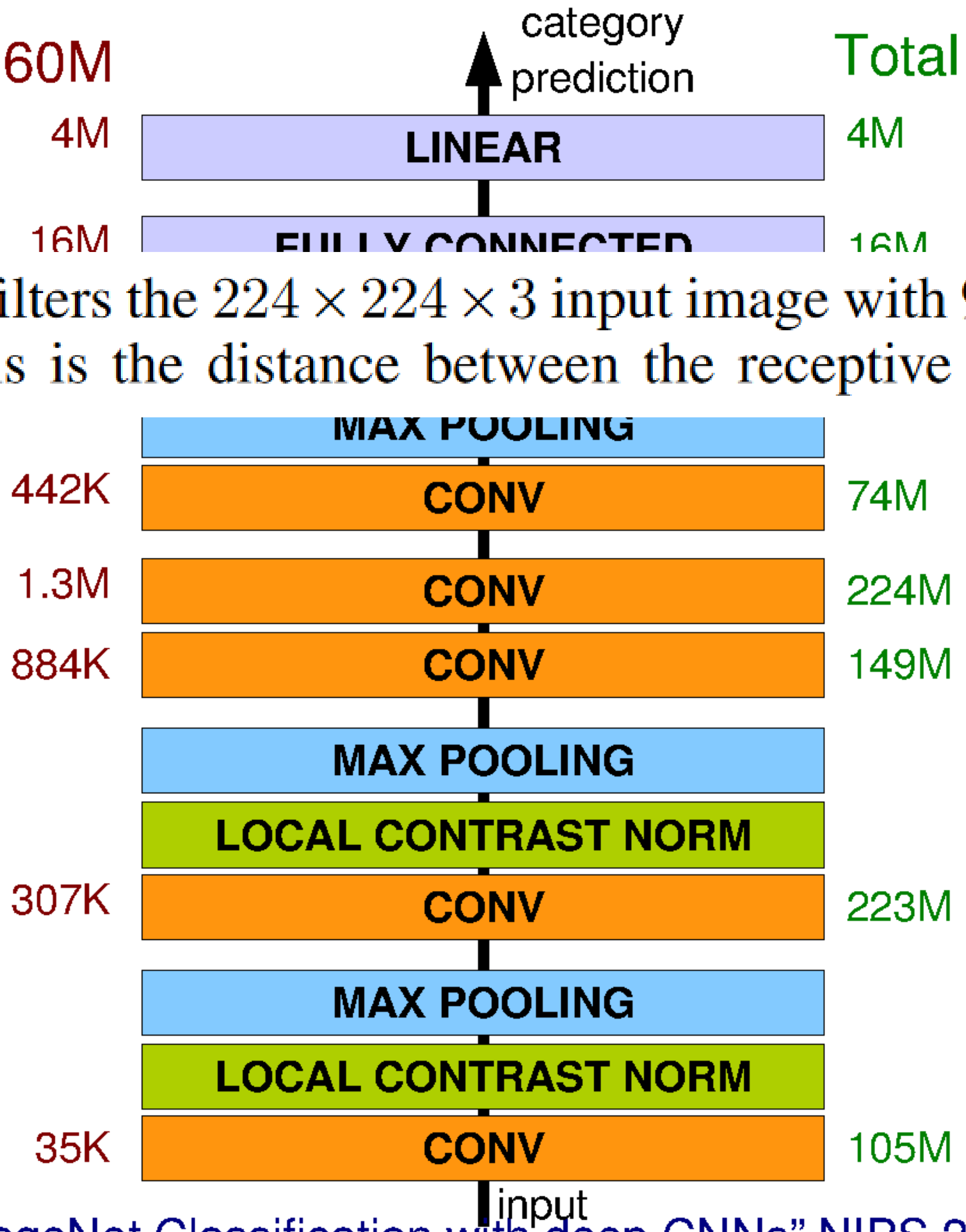


Krizhevsky et al. "ImageNet Classification with deep CNNs" NIPS 2012

Architecture for Classification

Total nr. params: 60M

Total nr. flops: 832M



The first convolutional layer filters the $224 \times 224 \times 3$ input image with 96 kernels of size $11 \times 11 \times 3$ with a stride of 4 pixels (this is the distance between the receptive field centers of neighboring

Krizhevsky et al. "ImageNet Classification with deep CNNs" NIPS 2012

Ranzato

Optimization

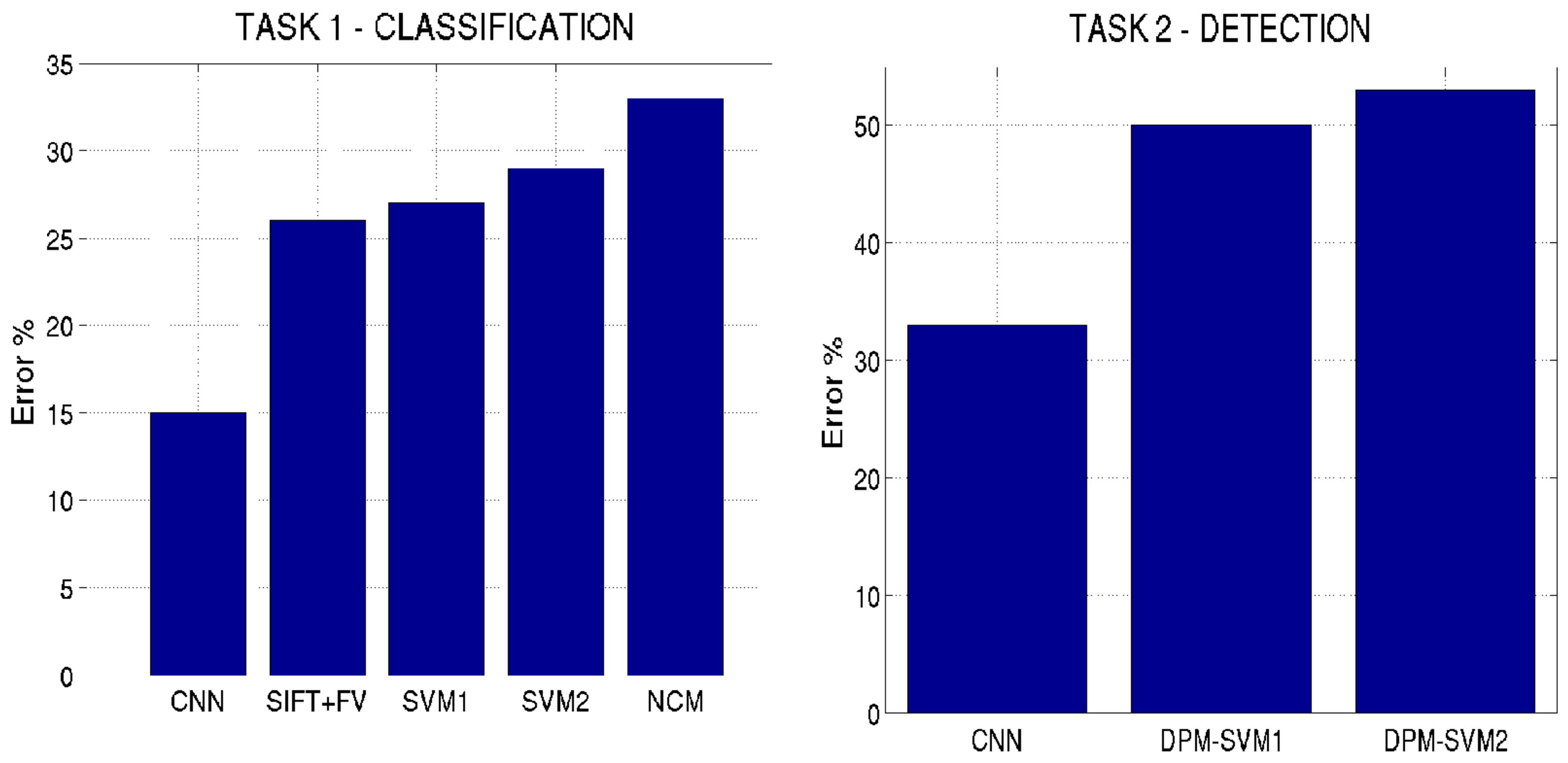
SGD with momentum:

- Learning rate = 0.01
- Momentum = 0.9

Improving generalization by:

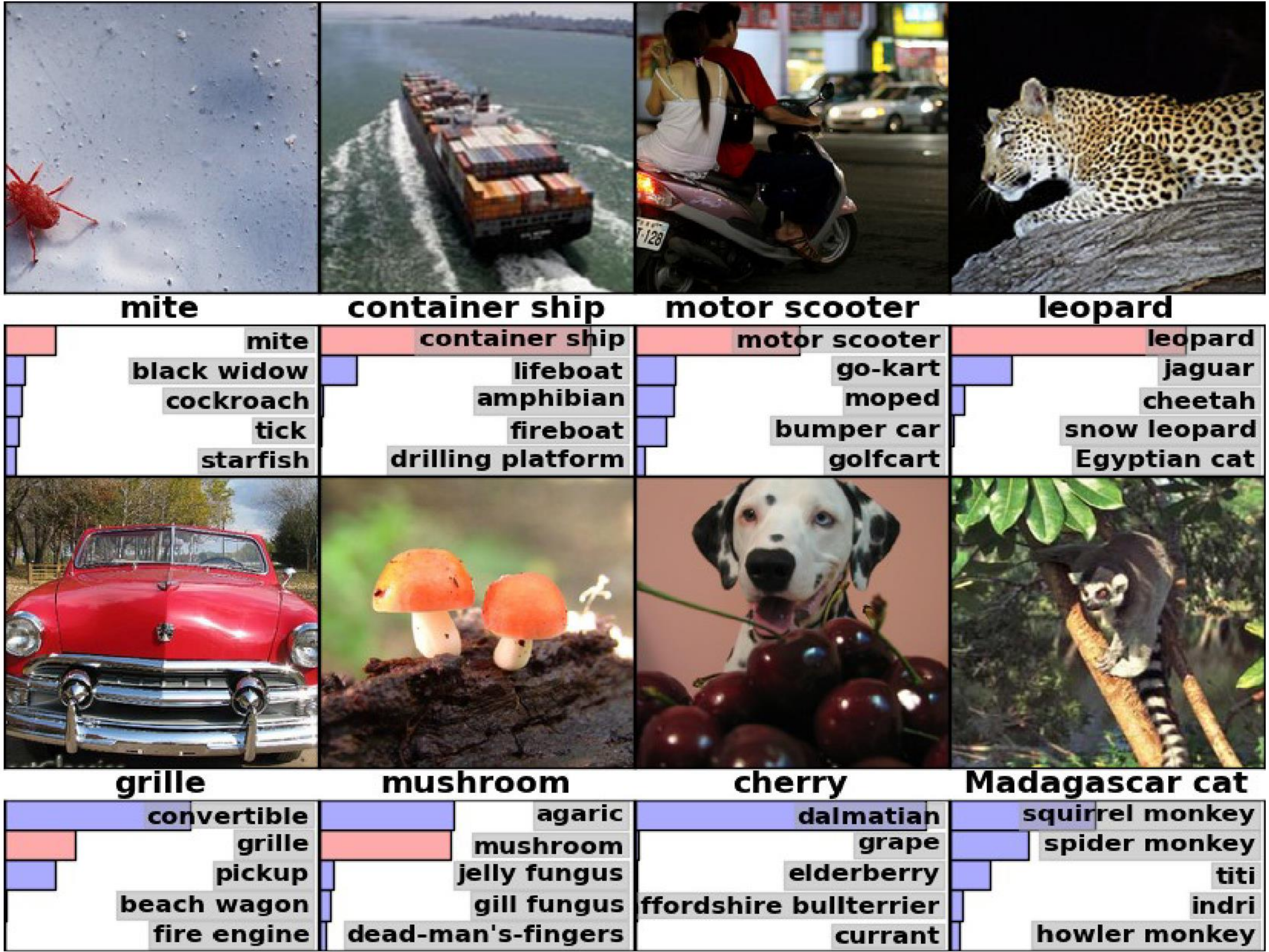
- Weight sharing (convolution)
- Input distortions
- Dropout = 0.5
- Weight decay = 0.0005

Results: ILSVRC 2012



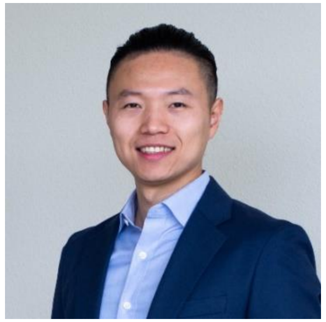
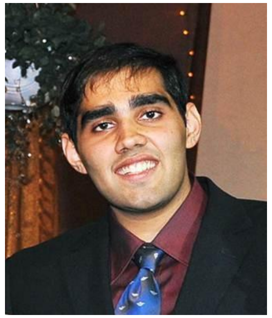
Krizhevsky et al. “ImageNet Classification with deep CNNs” NIPS 2012

Ranzato


mite
mite
black widow
cockroach
tick
starfish
container ship
container ship
lifeboat
amphibian
fireboat
drilling platform
motor scooter
motor scooter
go-kart
moped
bumper car
golfcart
leopard
leopard
jaguar
cheetah
snow leopard
Egyptian cat
grille
convertible
grille
pickup
beach wagon
fire engine
mushroom
agaric
mushroom
jelly fungus
gill fungus
dead-man's-fingers
cherry
dalmatian
grape
elderberry
ffordshire bullterrier
currant
Madagascar cat
squirrel monkey
spider monkey
titi
indri
howler monkey

Object Detectors Emerge in Deep Scene CNNs

Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba


Massachusetts Institute of Technology

CNN for Object Recognition

Large-scale image classification result on ImageNet

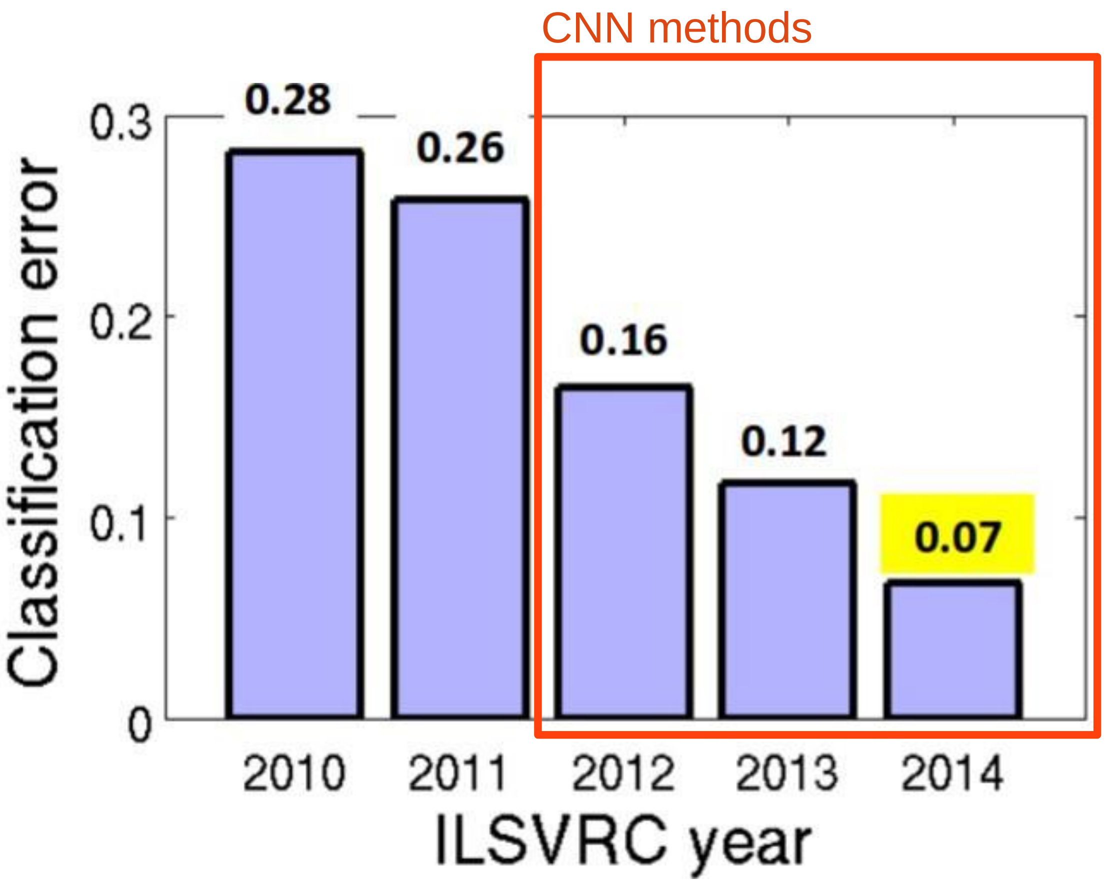
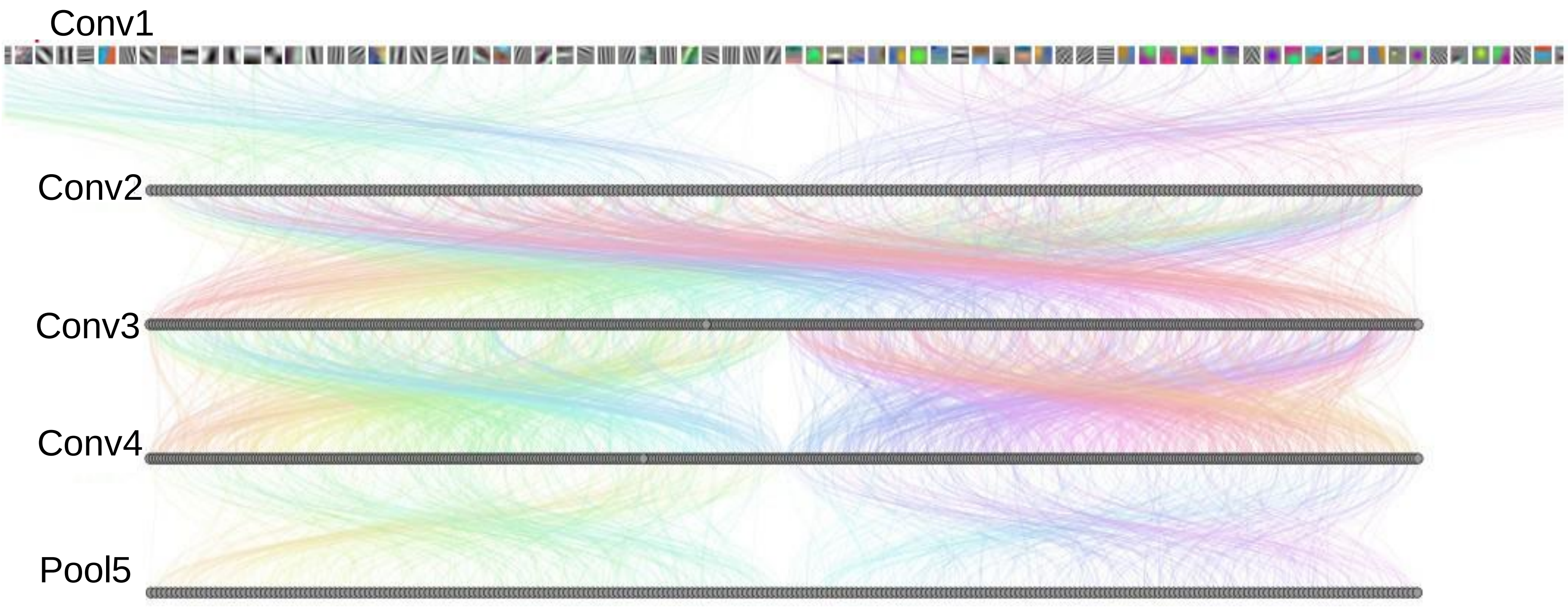


Figure from Olga Russakovsky ECCV'14 workshop

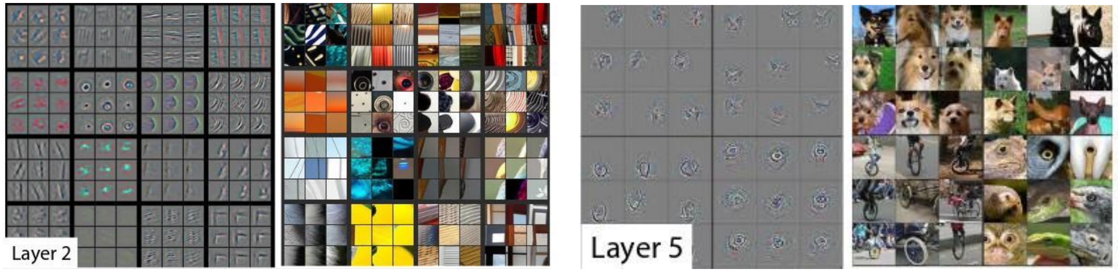
How Objects are Represented in CNN?



DrawCNN: visualizing the units' connections

How Objects are Represented in CNN?

Deconvolution

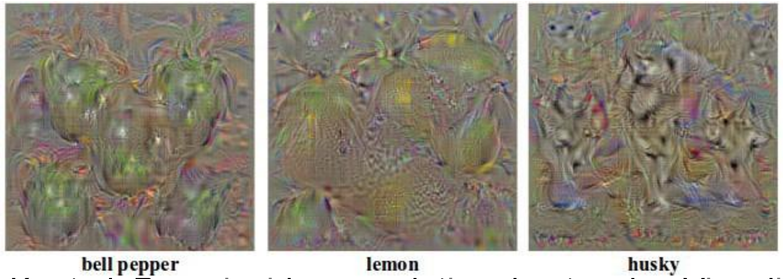


Zeiler, M. et al. Visualizing and Understanding Convolutional Networks,ECCV 2014.

Strong activation image

Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accu-rate object detection and semantic segmentation. CVPR 2014

Back-propagation



Simonyan, K. et al. Deep inside convolutional networks: Visualising image classification models and saliency maps. ICLR workshop, 2014

Another CNN interpretation method: Simplifying Scenes While Maintaining Classifier Decision

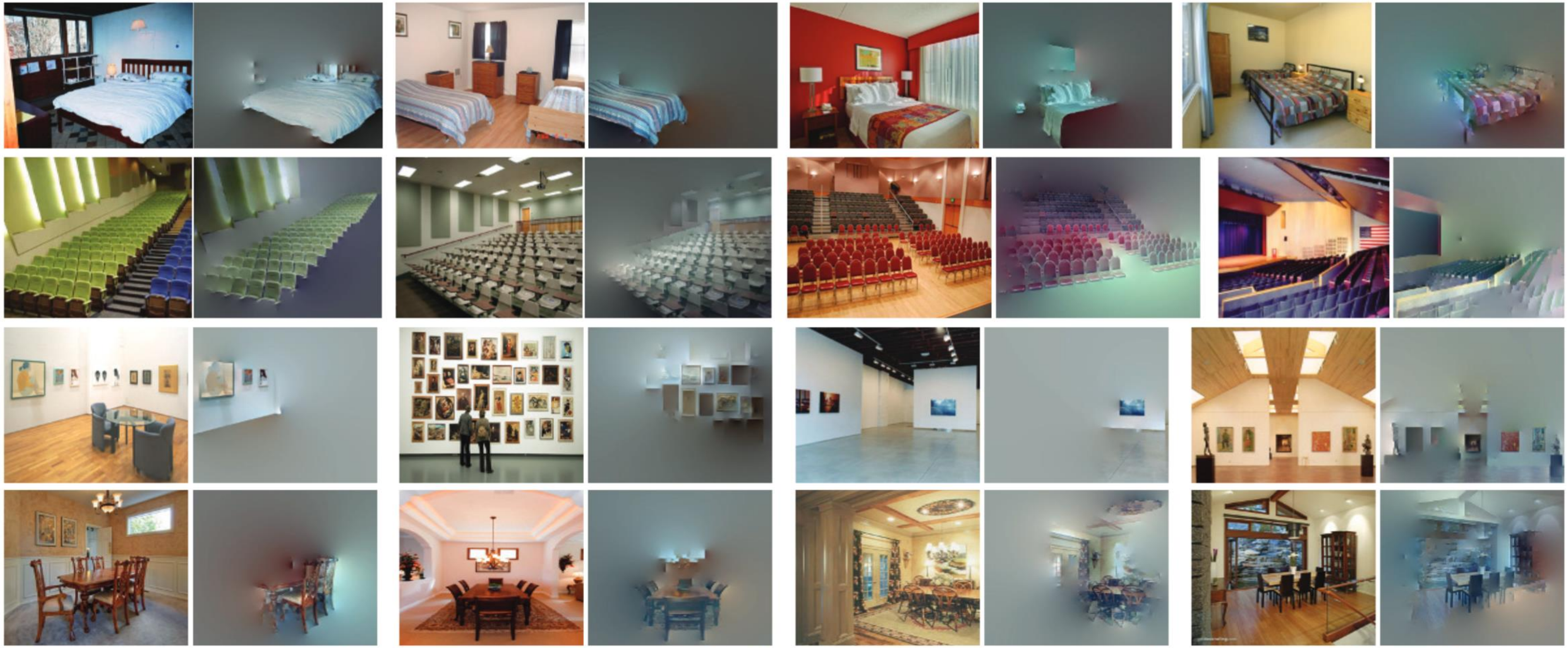


Figure 2: Each pair of images shows the original image (left) and a simplified image (right) that gets classified by the Places-CNN as the same scene category as the original image. From top to bottom, the four rows show different scene categories: bedroom, auditorium, art gallery, and dining room.

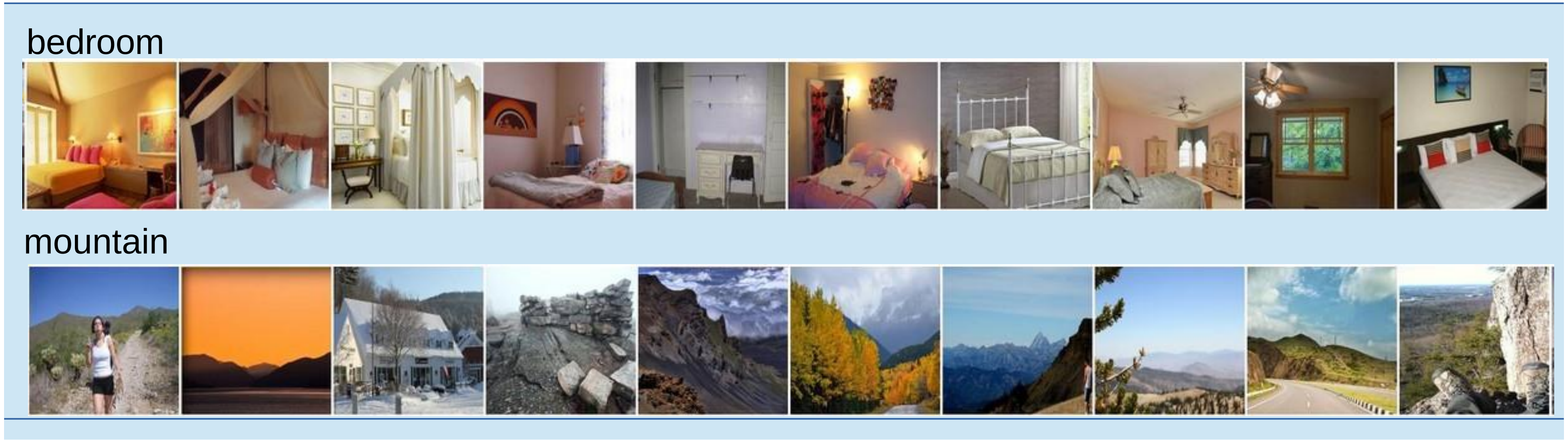
Another recognition task: Scene Recognition

Given an image, predict which place we are in.

Bedroom

Harbor

Learning to Recognize Scenes

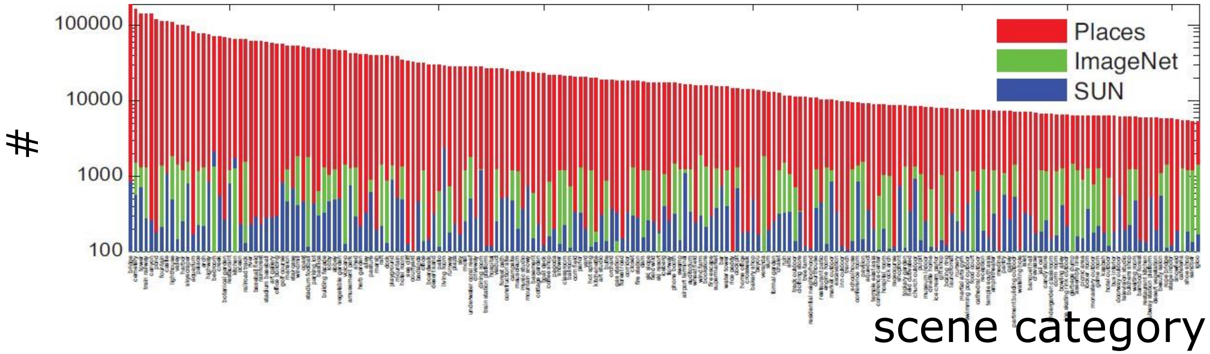


Possible internal representations:

- Objects (scene parts?)
- Scene attributes
- Object parts
- Textures

CNN for Scene Recognition

Places Database: 7 million images from 400 scene categories



Places-CNN: AlexNet CNN on 2.5 million images from 205 scene categories.

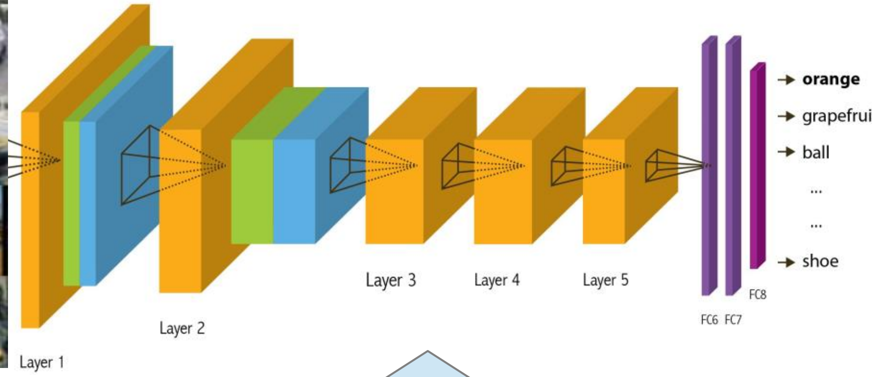
	Places 205	SUN 205
Places-CNN	**50.0%**	**66.2%**
ImageNet CNN feature+SVM	40.8%	49.6%

Zhou, et al. NIPS, 2014.

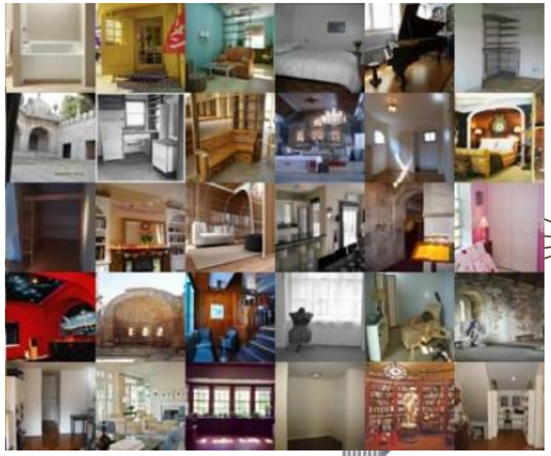
ImageNet CNN and Places CNN

IMAGENET

ImageNet CNN for Object Classification

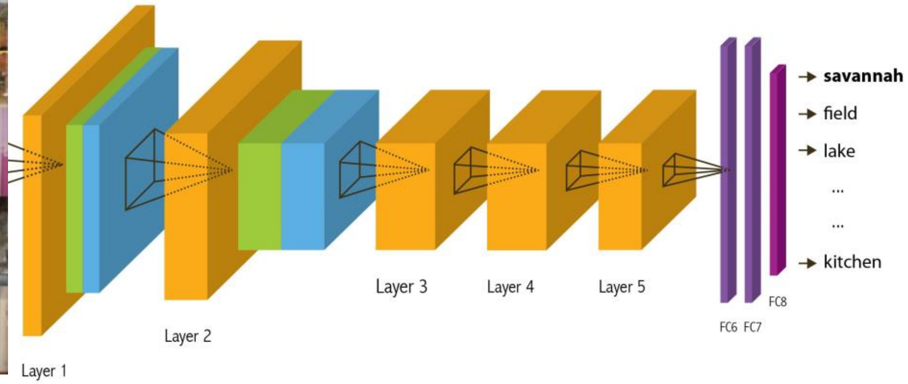


Same architecture: AlexNet



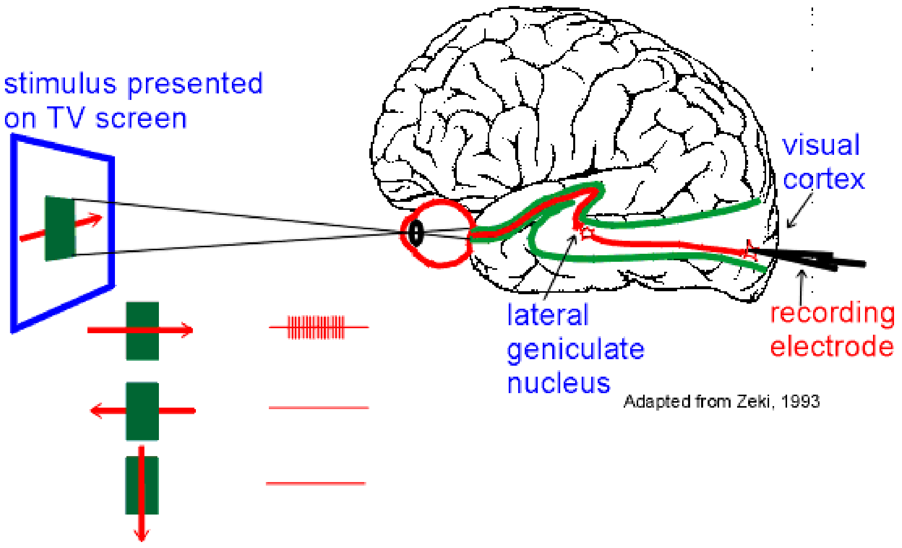
Places

Places CNN for Scene Classification

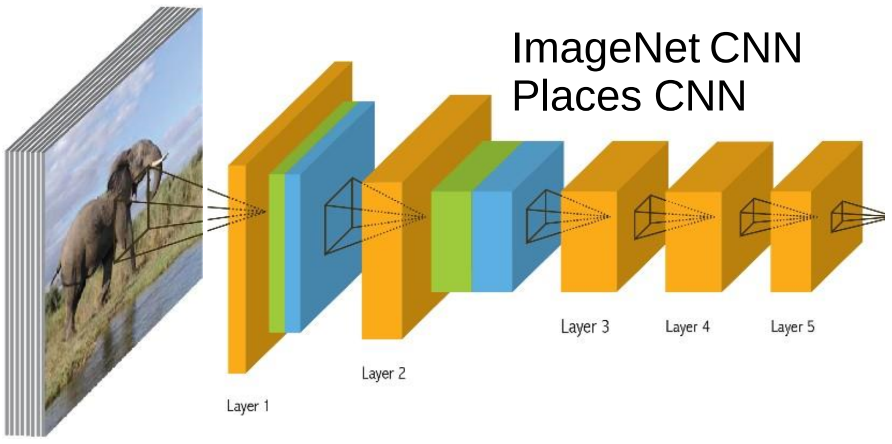


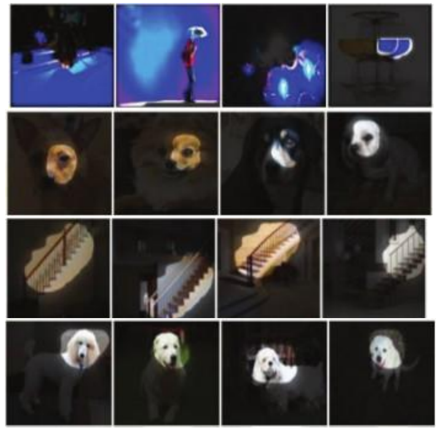
Data-Driven Approach to Study CNN

Neuroscientists study brain

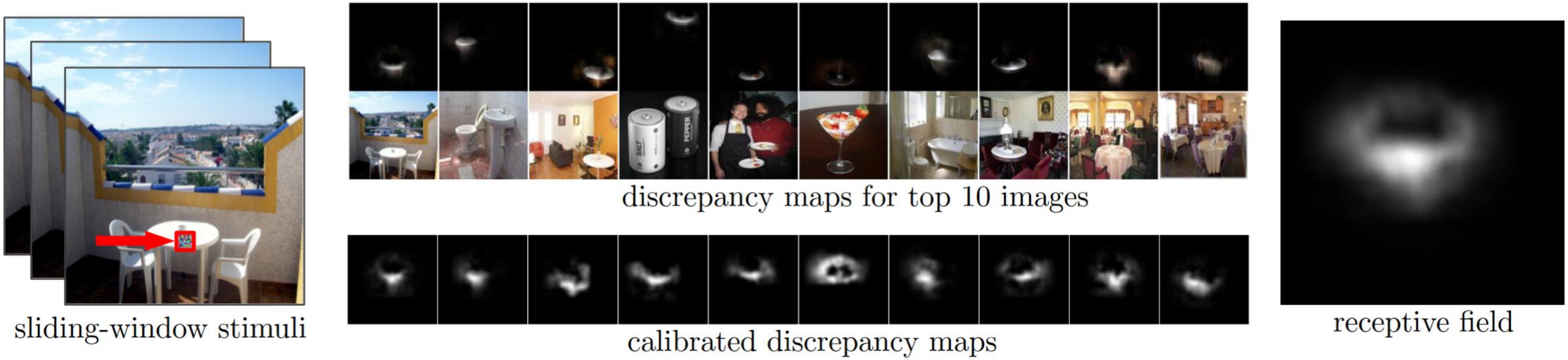


200,000 image stimuli of objects and scene categories (ImageNet TestSet+SUN database)





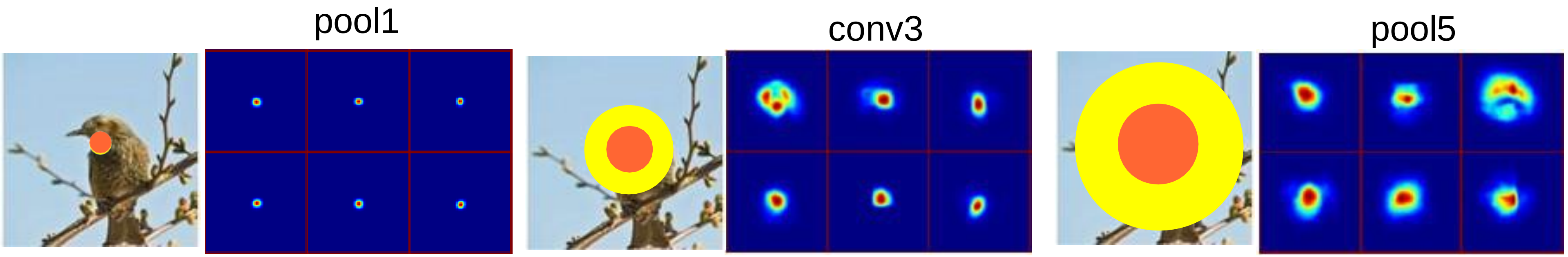
Estimating the Receptive Fields



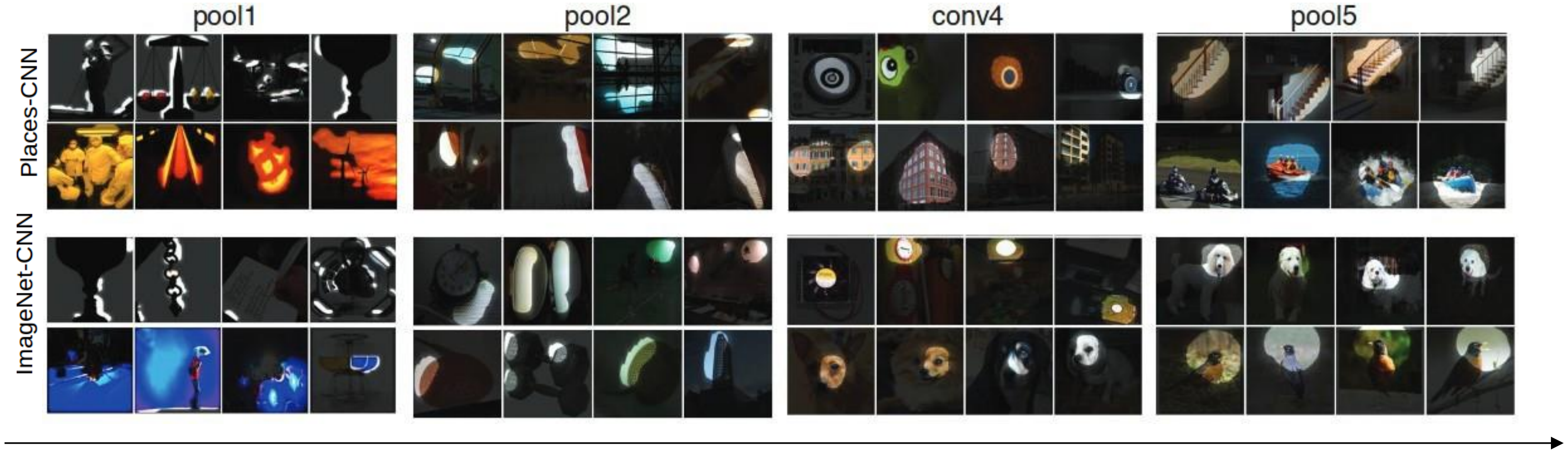
Estimating the Receptive Fields

Estimated receptive fields

Actual size of RF is much smaller than the theoretic size



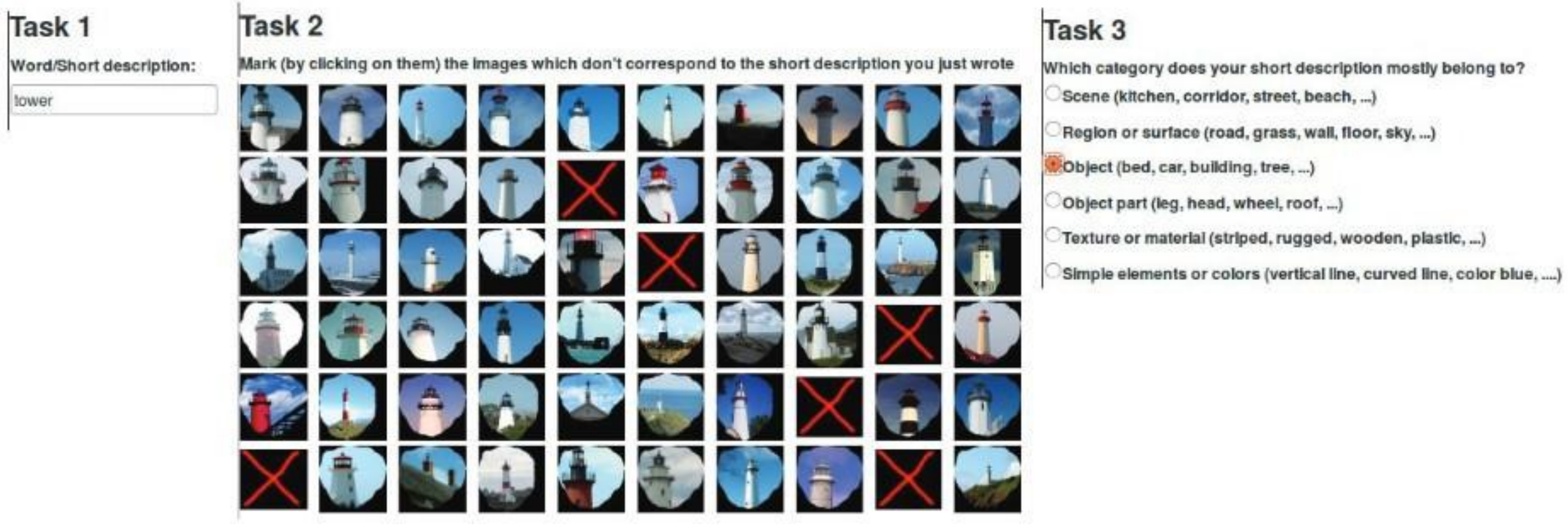
Segmentation using the RF of Units



More semantically meaningful

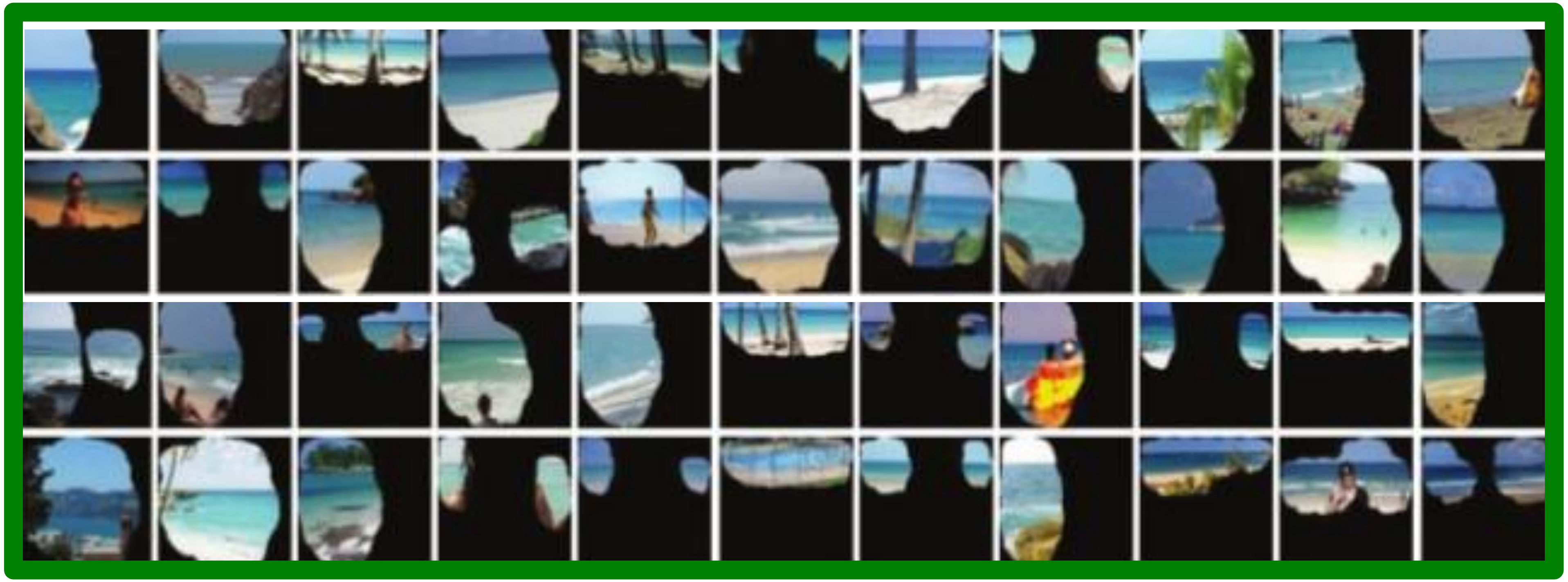
Annotating the Semantics of Units

Top ranked segmented images are cropped and sent to Amazon Turk for annotation.



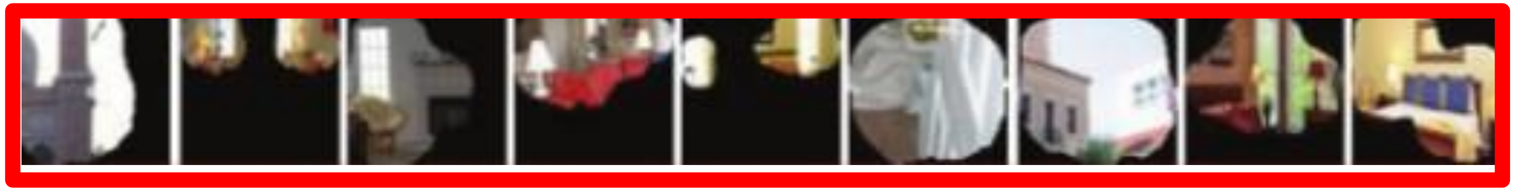
Annotating the Semantics of Units

Pool5, unit 76; Label: ocean; Type: scene; Precision: 93%



Annotating the Semantics of Units

Pool5, unit 13; Label: Lamps; Type: object; Precision: 84%

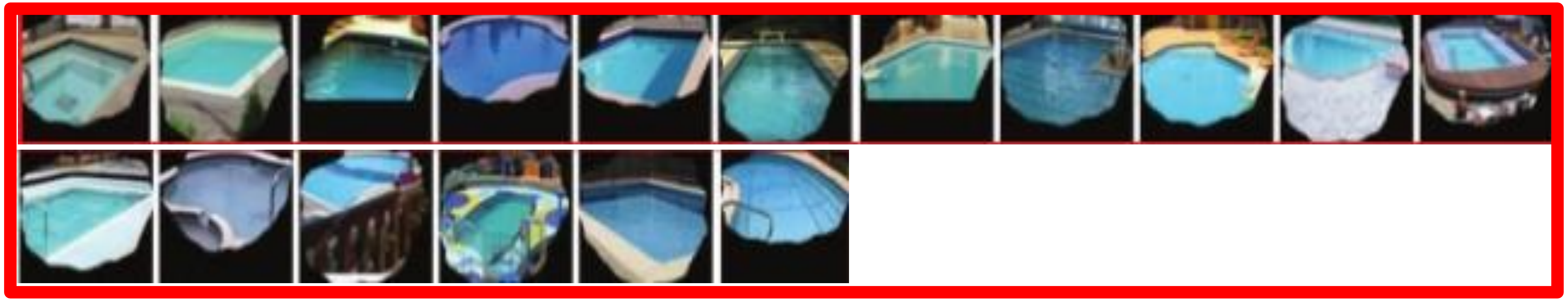


Annotating the Semantics of Units

Pool5, unit 77; Label:legs; Type: object part; Precision: 96%

Annotating the Semantics of Units

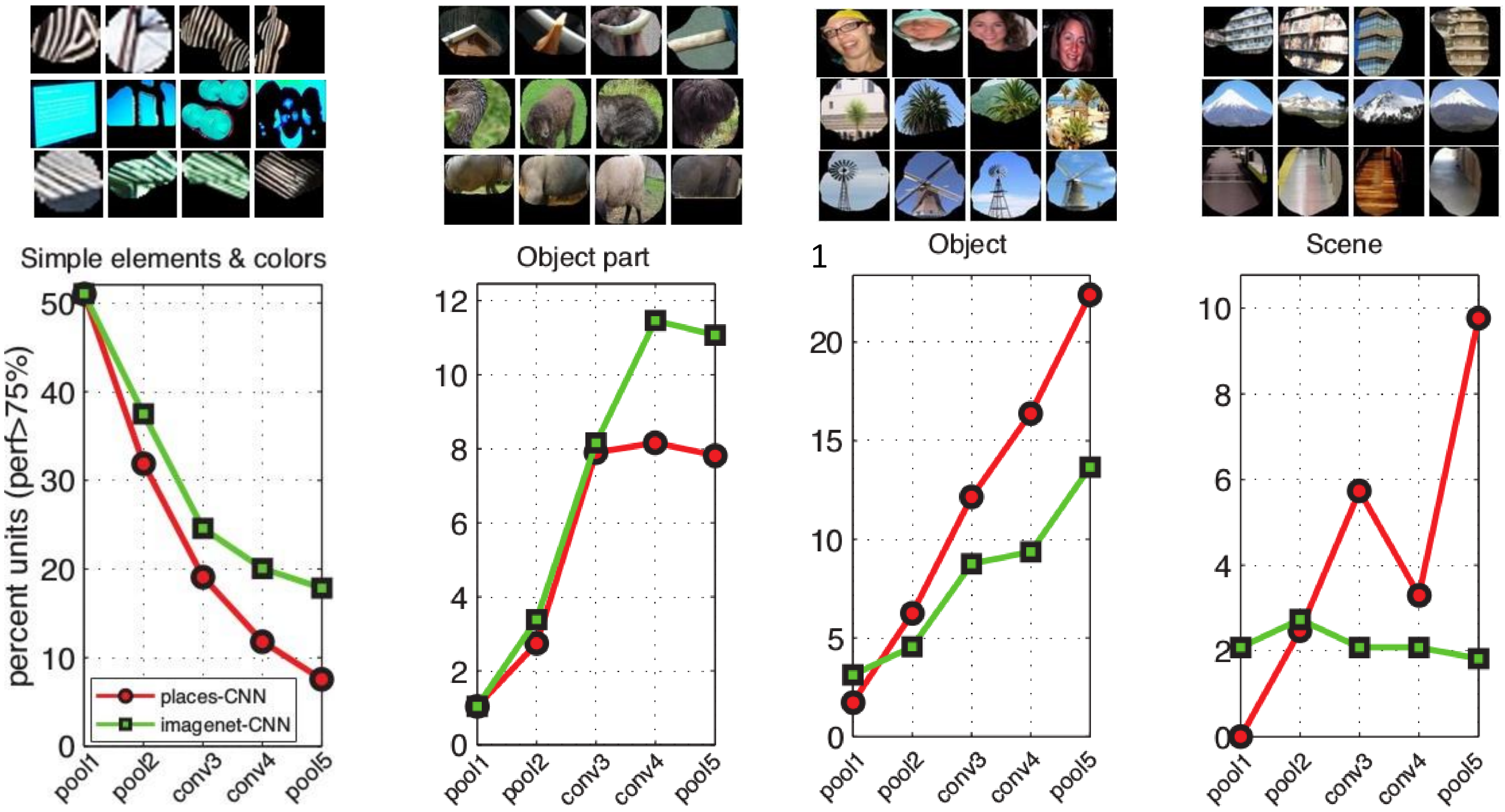
Pool5, unit 112; Label: pool table; Type: object; Precision: 70%



Annotating the Semantics of Units

Pool5, unit 22; Label: dinner table; Type: scene; Precision: 60%

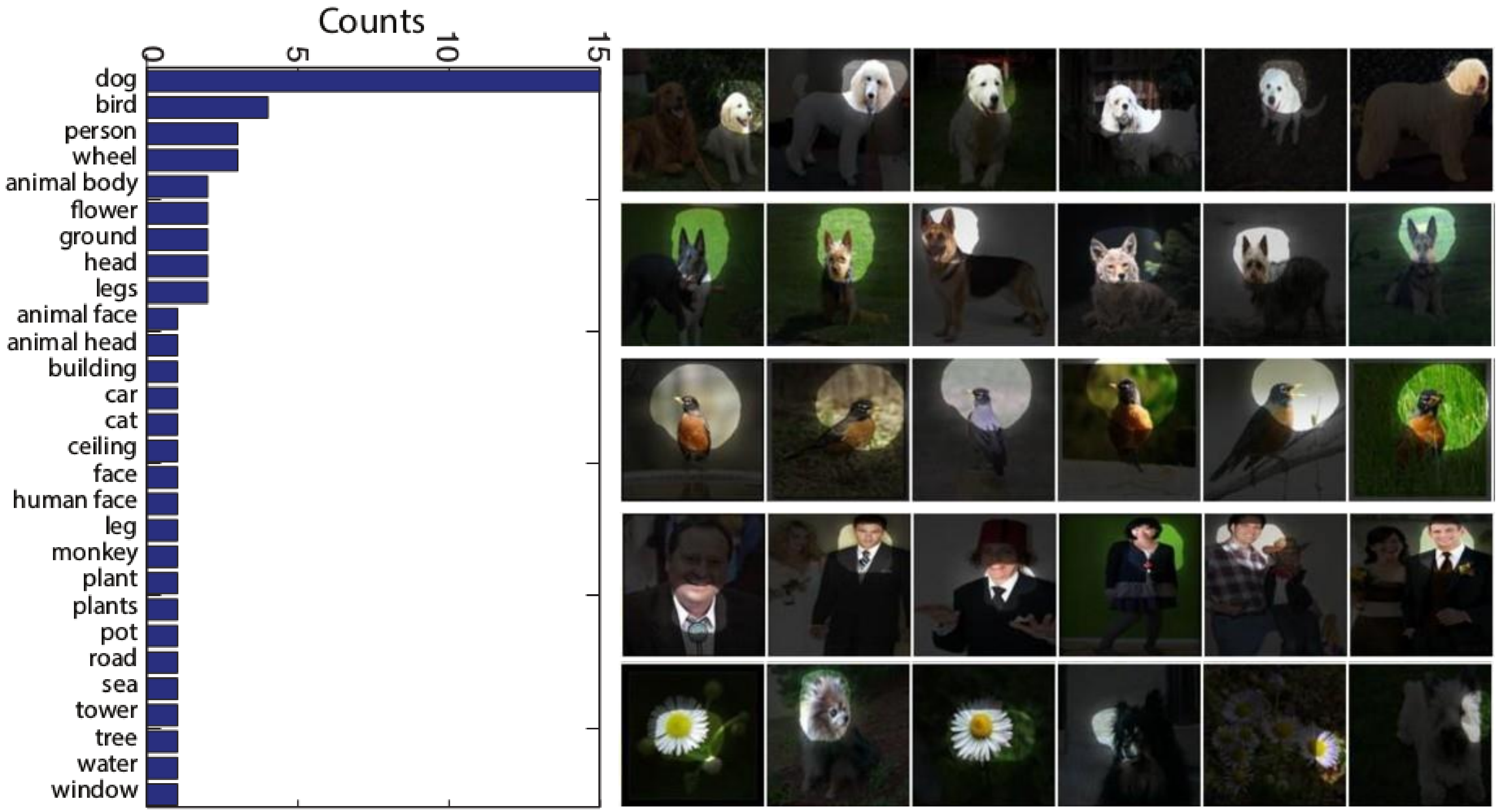
Distribution of Semantic Types at Each Layer



Object detectors emerge within CNN trained to classify scenes, without any object supervision!

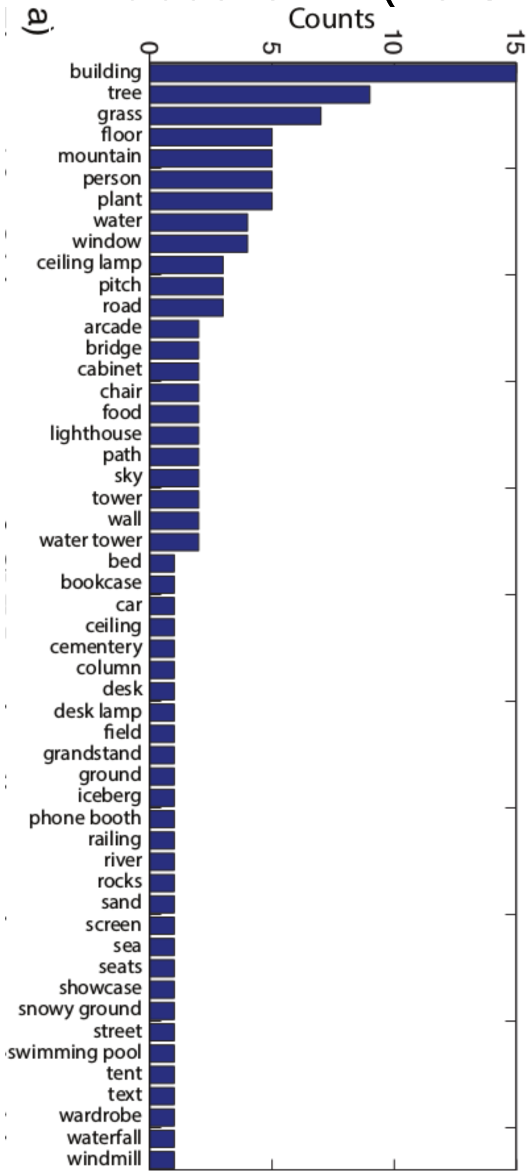
Histogram of Emerged Objects in Pool5

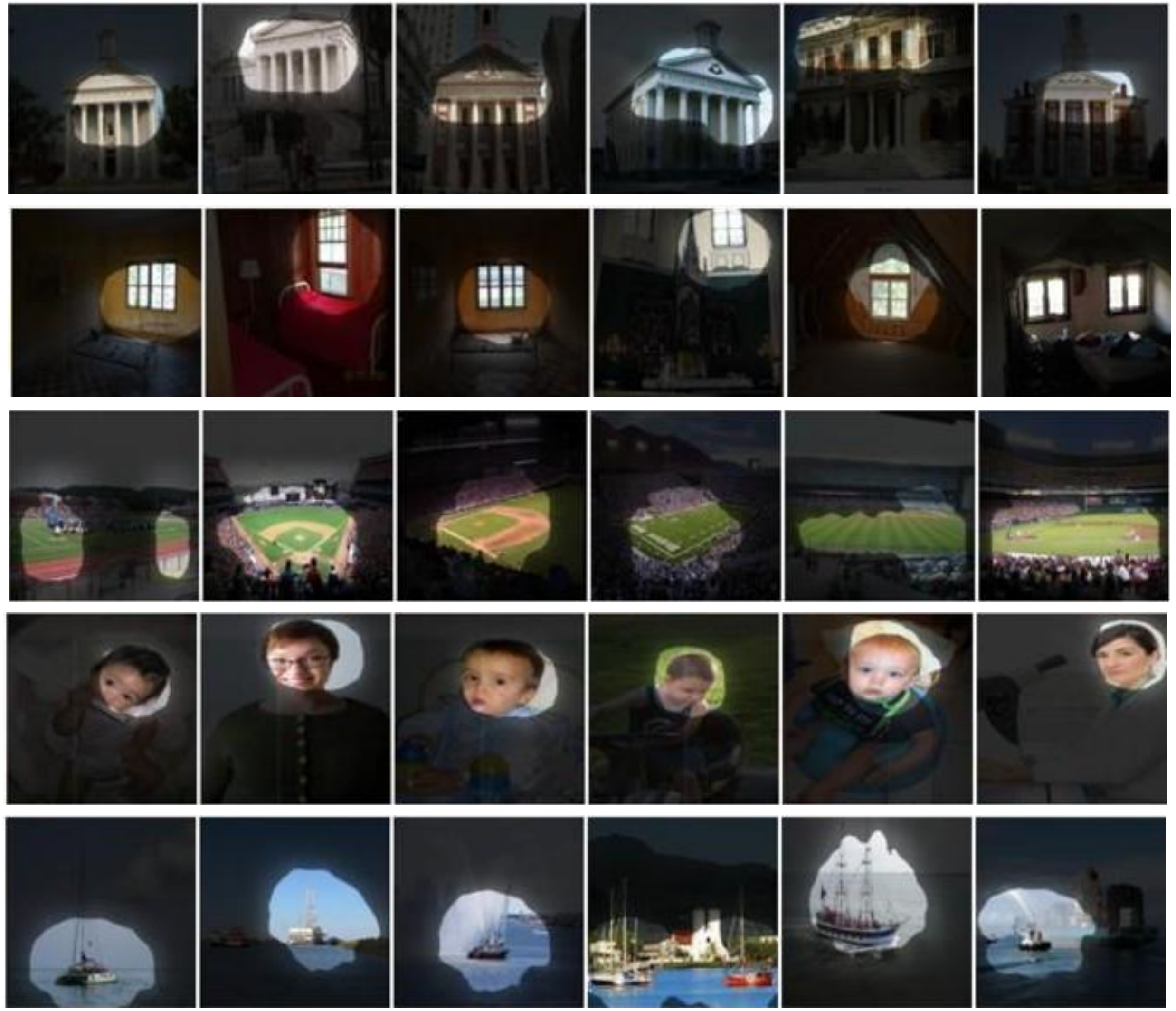
ImageNet-CNN (59/256)



Histogram of Emerged Objects in Pool5

Places-CNN (151/256)






Buildings
56) building
120) arcade
8) bridge
123) building
119) building
9) lighthouse
Furniture
18) billard table
155) bookcase
116) bed
38) cabinet
85) chair
People
3) person
49) person
138) person
100) person
Lighting
55) ceiling lamp
174) ceiling lamp
223) ceiling lamp
13) desk lamp
Nature
195) grass
89) iceberg
140) mountain
159) sand

Wrap up

- There are many ways to visualize what a neural network has learned
- Networks learn smaller receptive fields than the "theoretical" receptive field.
- As you go deeper in the network, the hidden activations correspond more to high-level semantic concepts
- Object detectors emerge inside a CNN trained to classify scenes, without any object supervision.