Data Mining & Machine Learning

CS37300
Purdue University

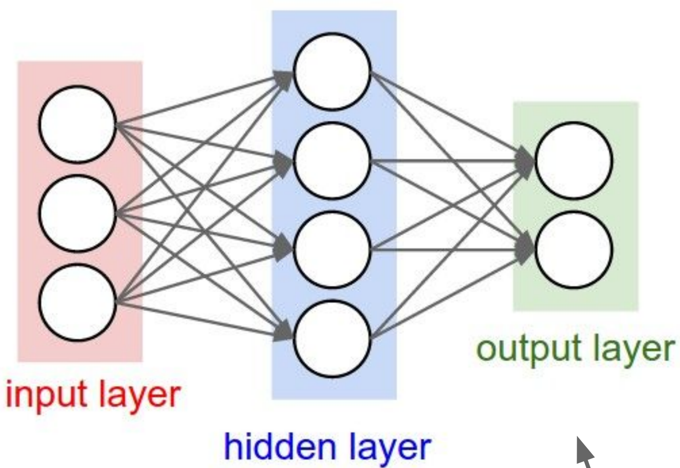
October 25, 2017

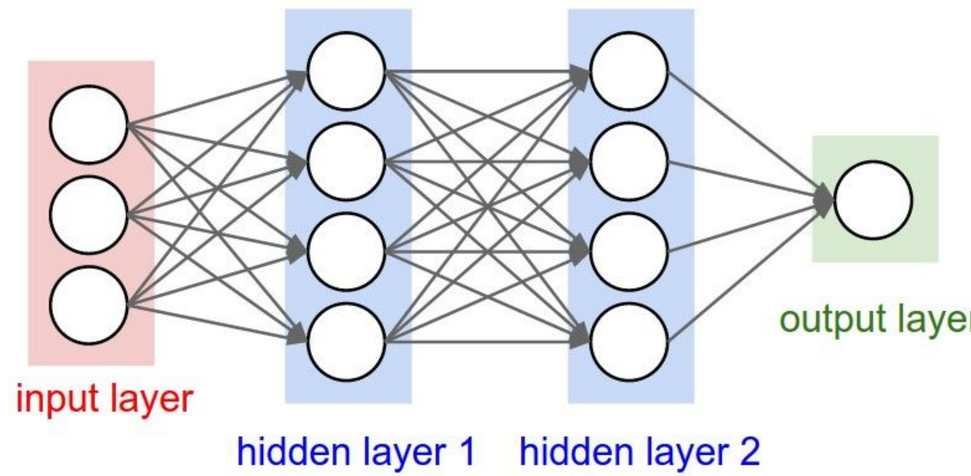
Deep Learning Architectures

Overview

- Deep Learning Architectures
    - Feedforward neural networks
    - Convolutional neural networks
        - Filters
        - Filter Banks
        - Max Pooling

Feedforward Neural Networks





"2-layer Neural Net", or
"1-hidden-layer Neural Net"

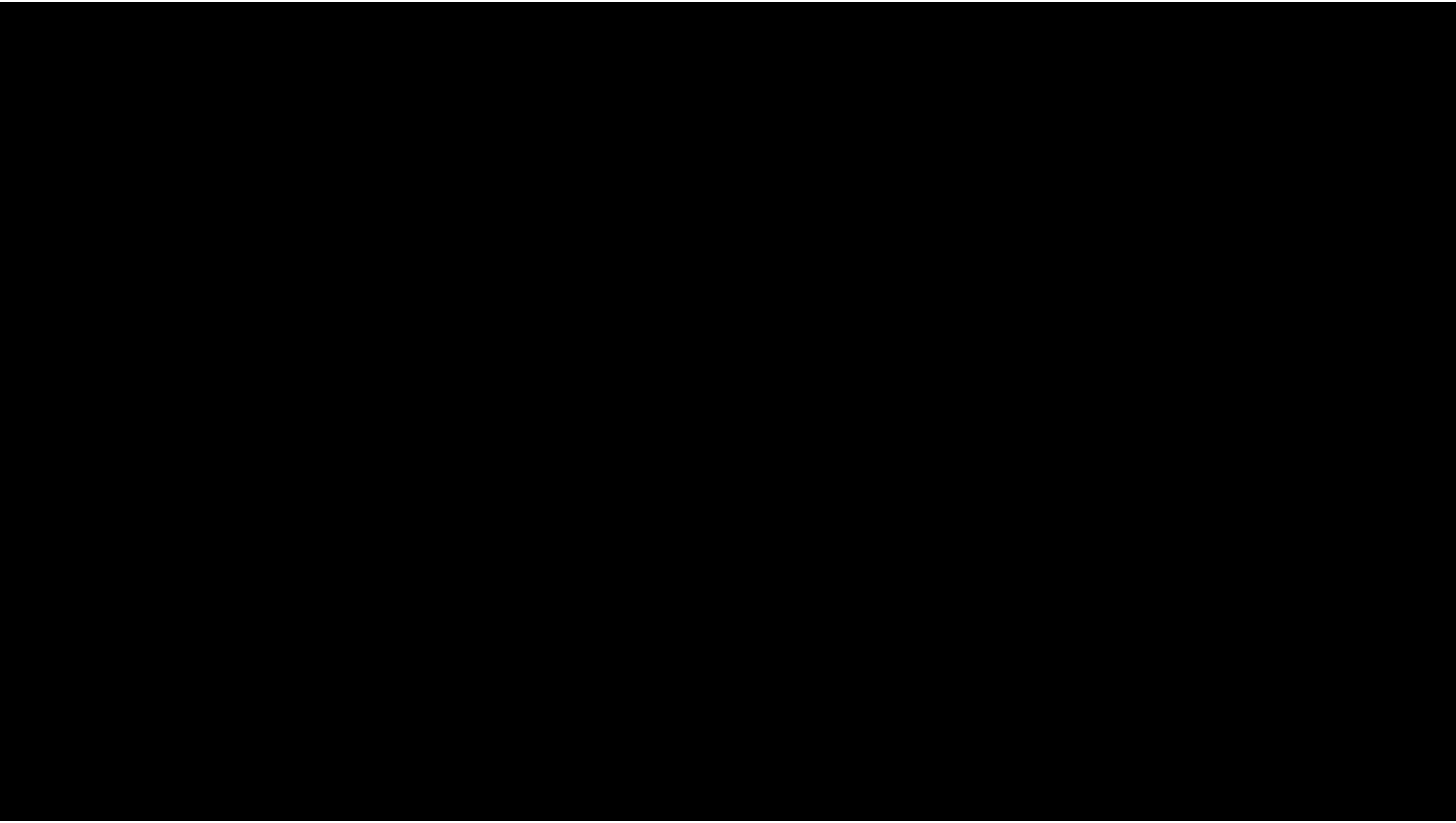
"Fully-connected" layers

"3-layer Neural Net", or
"2-hidden-layer Neural Net"

Ack: Fei-Fei Li & Andrej Karpathy & Justin Johnson

- Layers do not need to be fully connected
- Size of layer (number of units) is another parameter

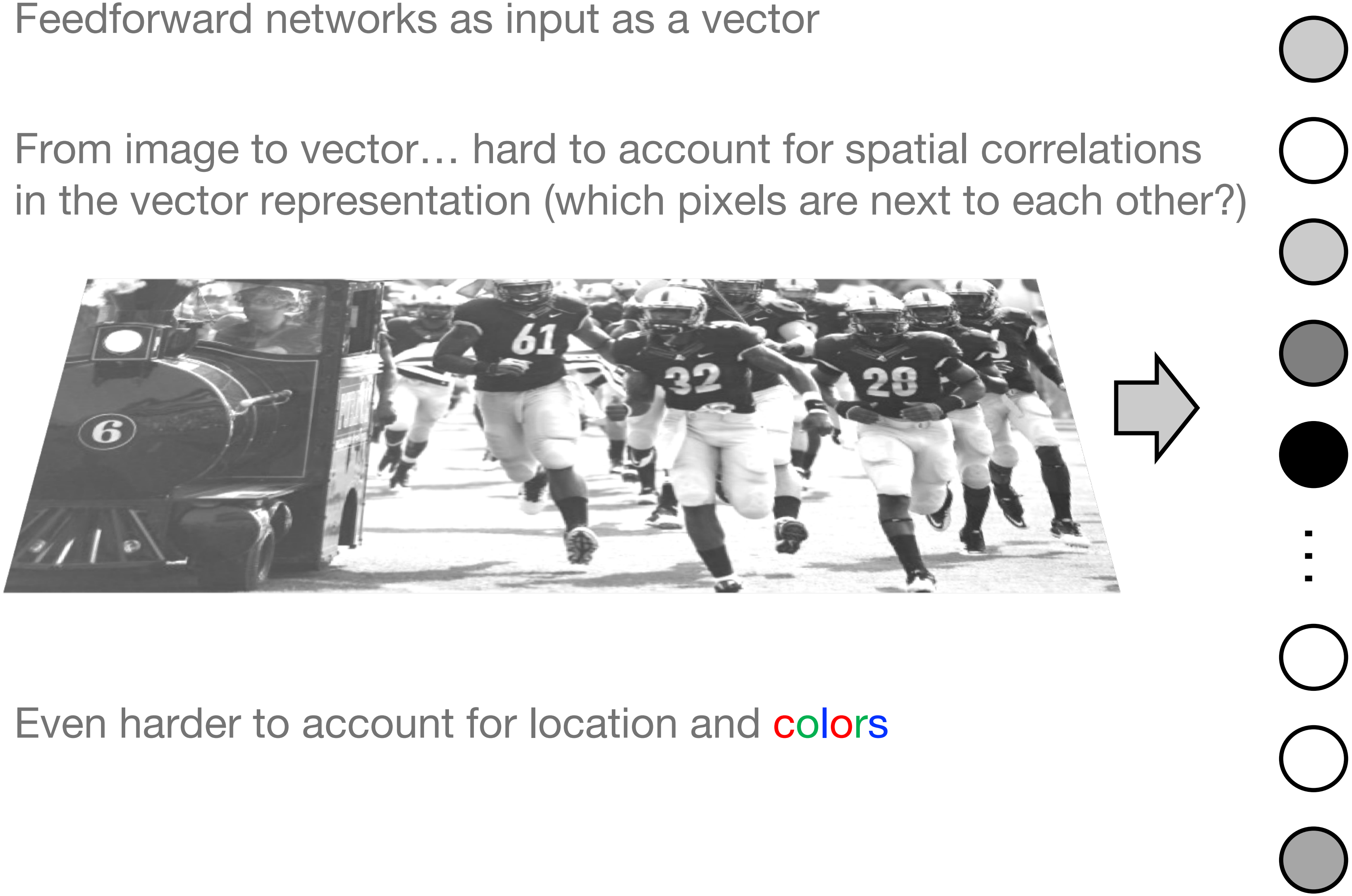
How Feedforward Networks Work



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

Convolutional Neural Networks (Motivation)

- Feedforward networks as input as a vector
- From image to vector… hard to account for spatial correlations in the vector representation (which pixels are next to each other?)



- Even harder to account for location and colors

Convolutional Neural Network (CNN)

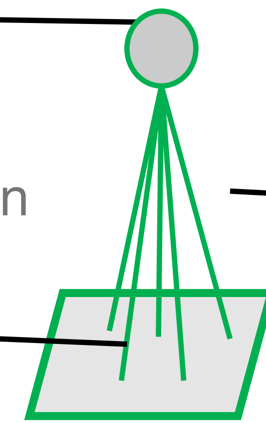
Non-linear activation

Transforms large image patch into one output

Basic unit: single neuron

single neuron

Image patch



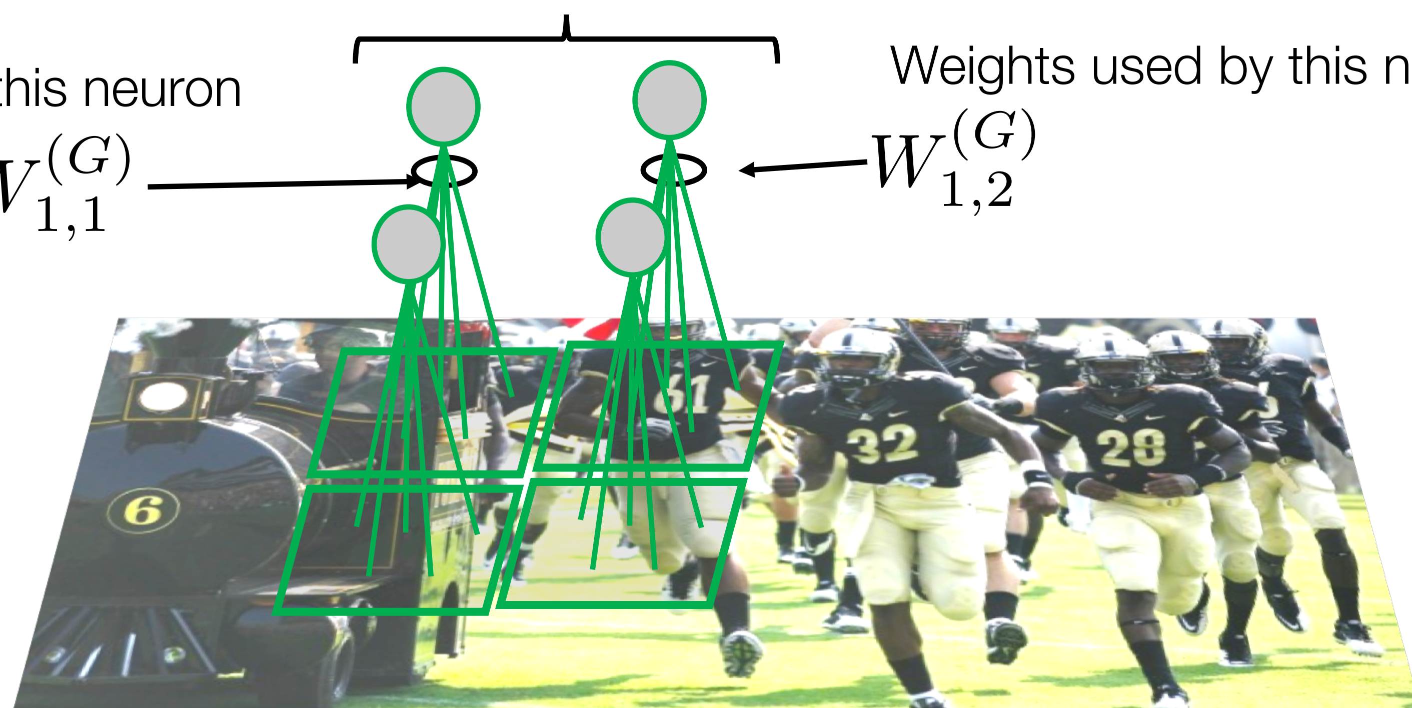
filter G

Weights used by this neuron

$W_{1,1}^{(G)}$

Weights used by this neuron

$W_{1,2}^{(G)}$



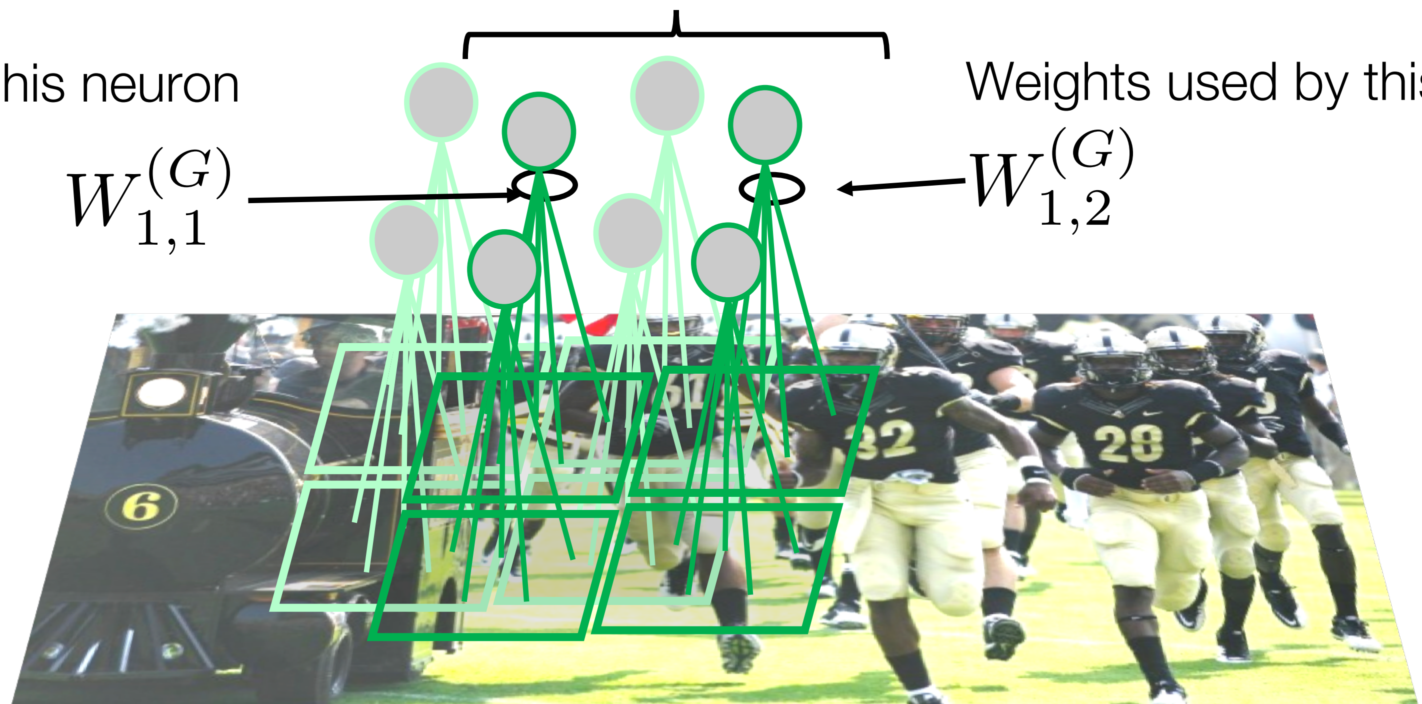
Convolutional Neural Network (CNN)

Cover the rest of the image by sliding the filter

filter G

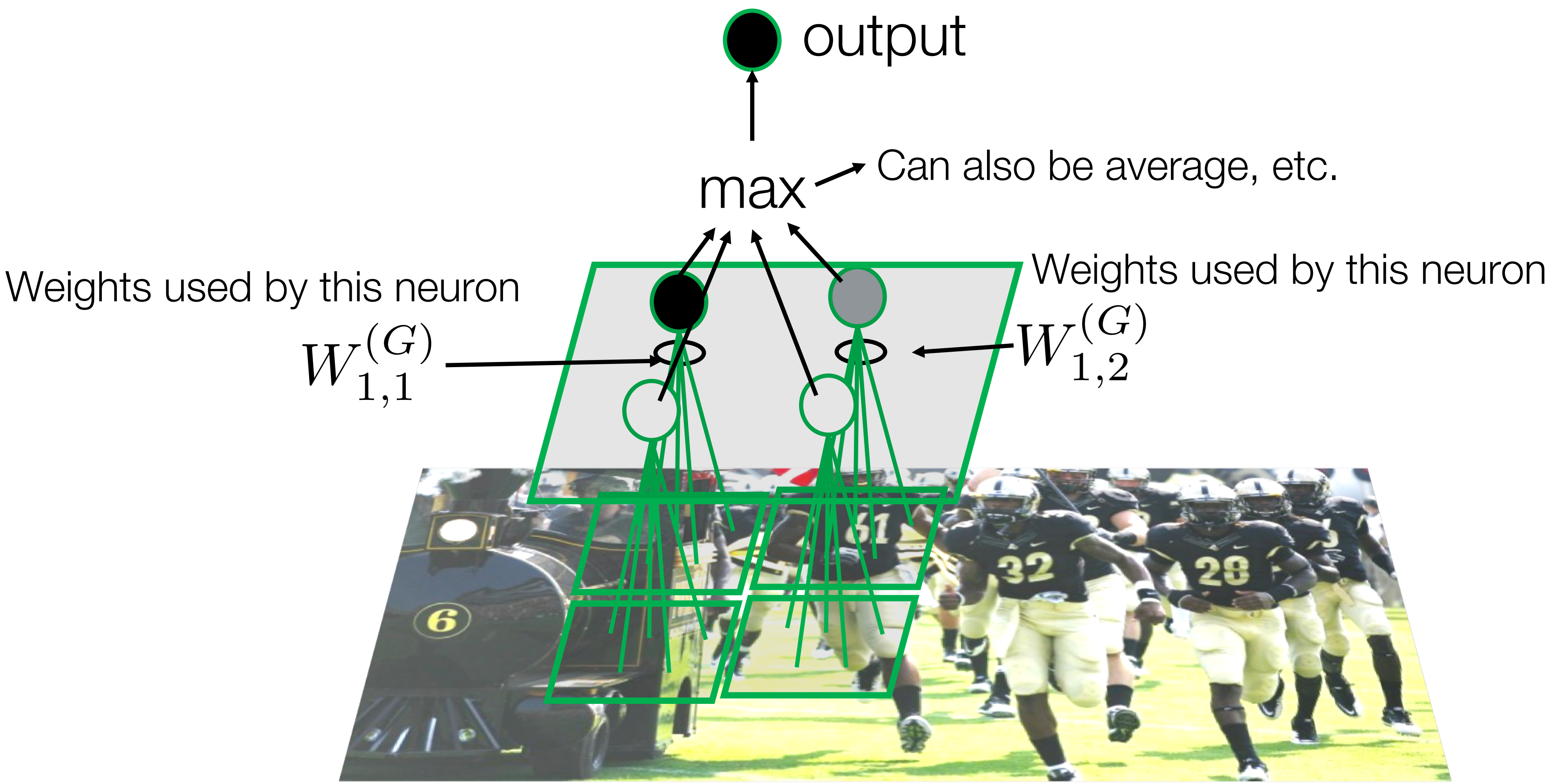
Weights used by this neuron $W^{(G)}_{1,1}$

Weights used by this neuron $W^{(G)}_{1,2}$



Pooling

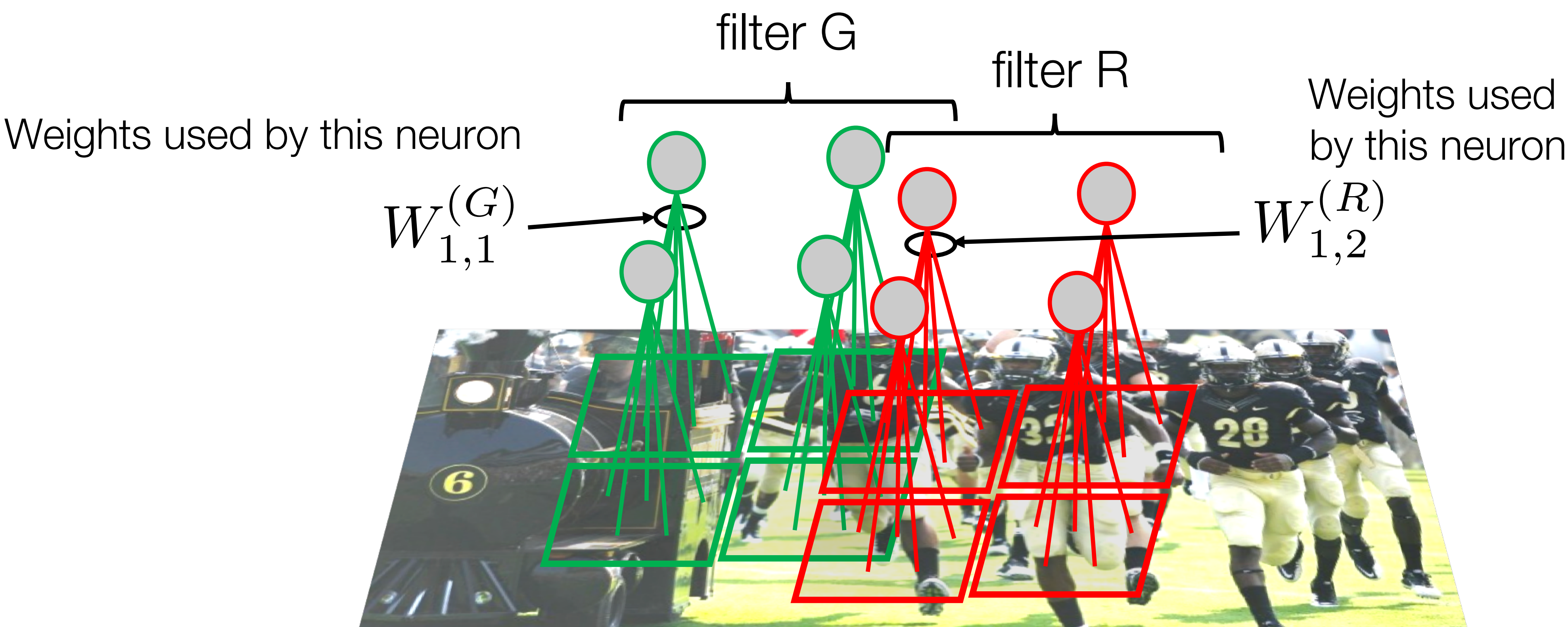
- Max pooling is a way to get a single output out of a filter



How to represent the image with different colors?

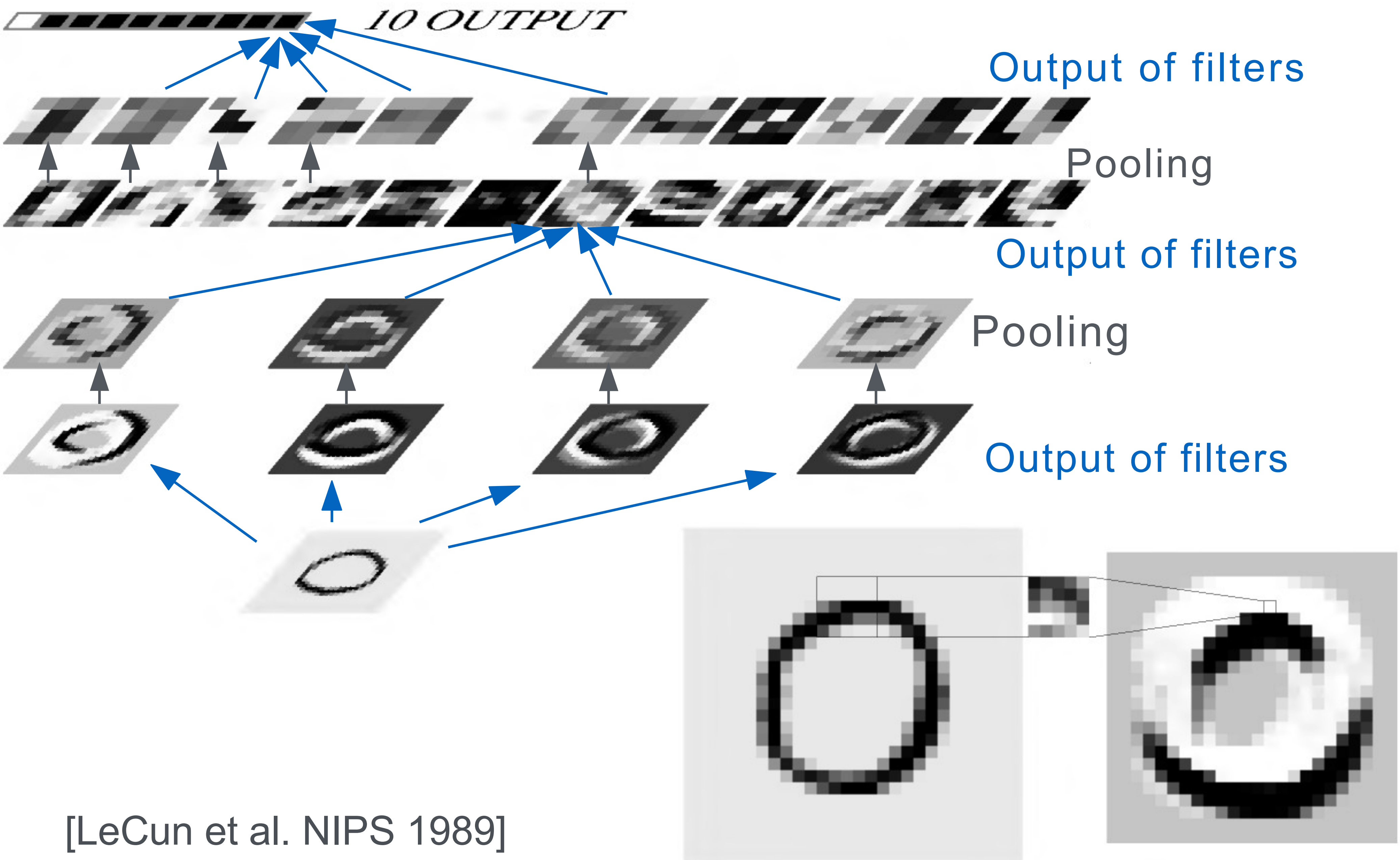
Convolutional Neural Network (CNN)

- We can use filters that focus on each color…



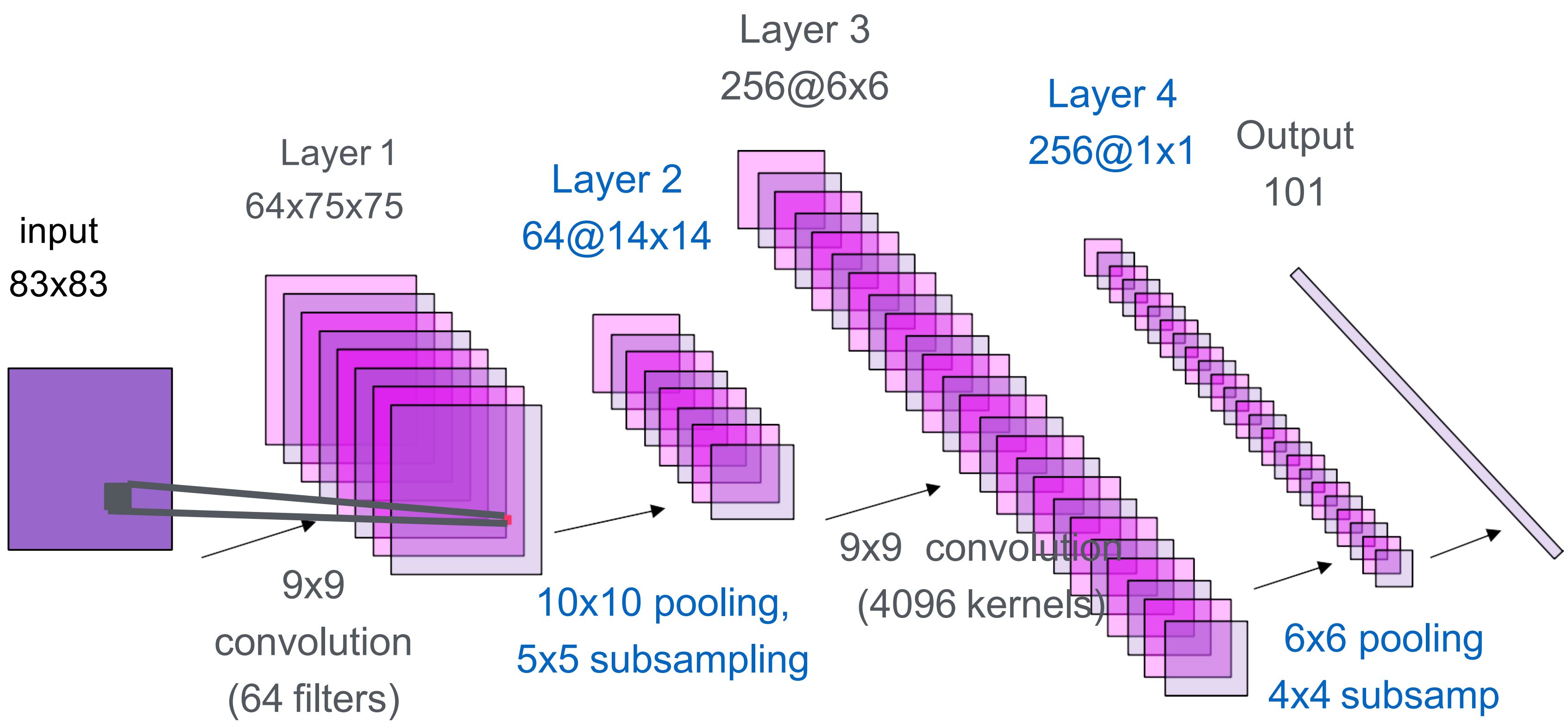
We can also have more than one filter for each color

Convolutional network



[LeCun et al. NIPS 1989]

Convolutional Network (ConvNet)

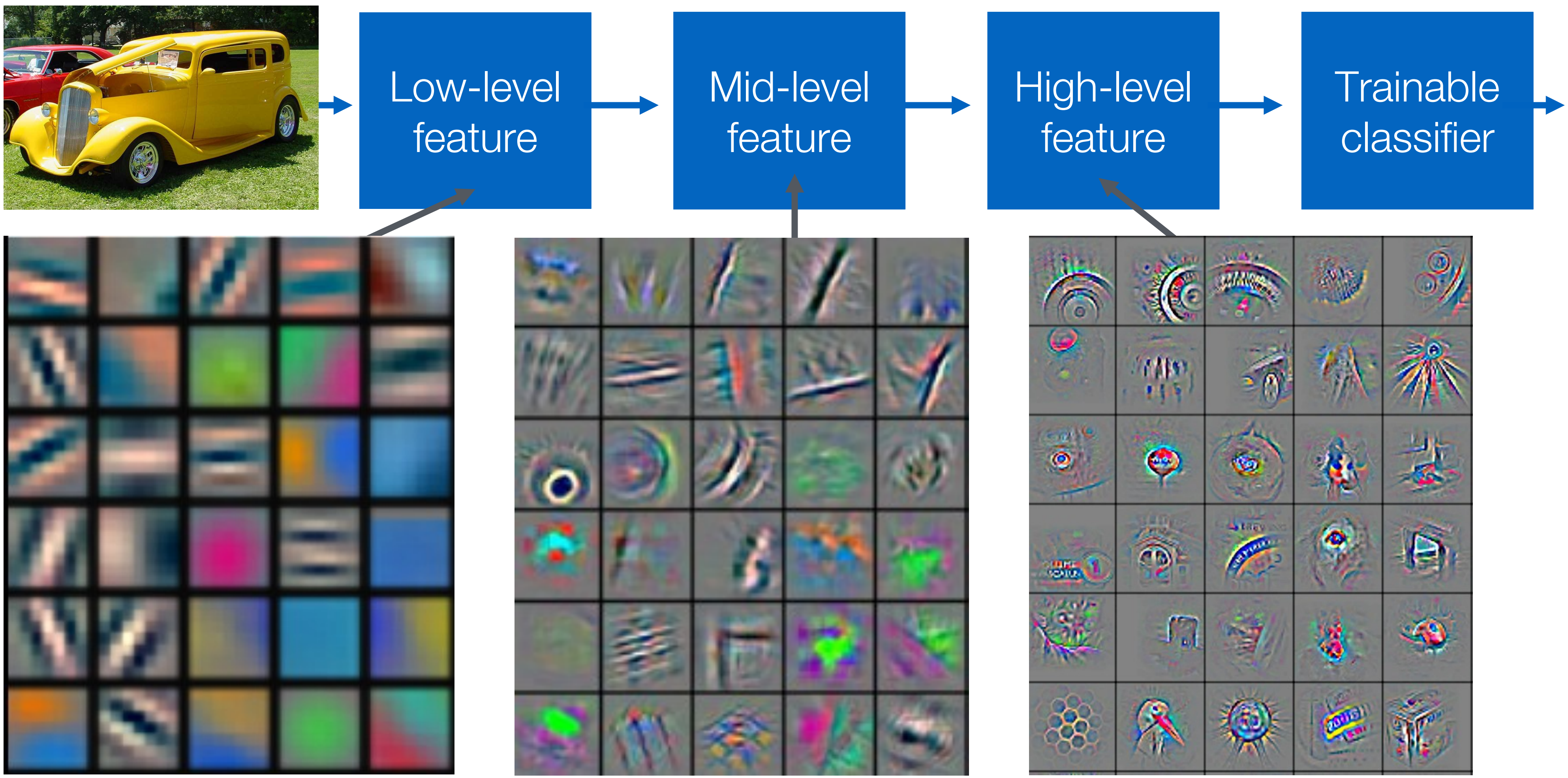


– Non-Linearity: sigmoid, half-wave rectification (ReLU)
– Pooling: max, average,…
– Training: Image labels

Original: Y. LeCun, Facebook, modified by Ribeiro

Deep learning = learning hierarchical representations

It's deep if it has more than one stage of non-linear feature transformation



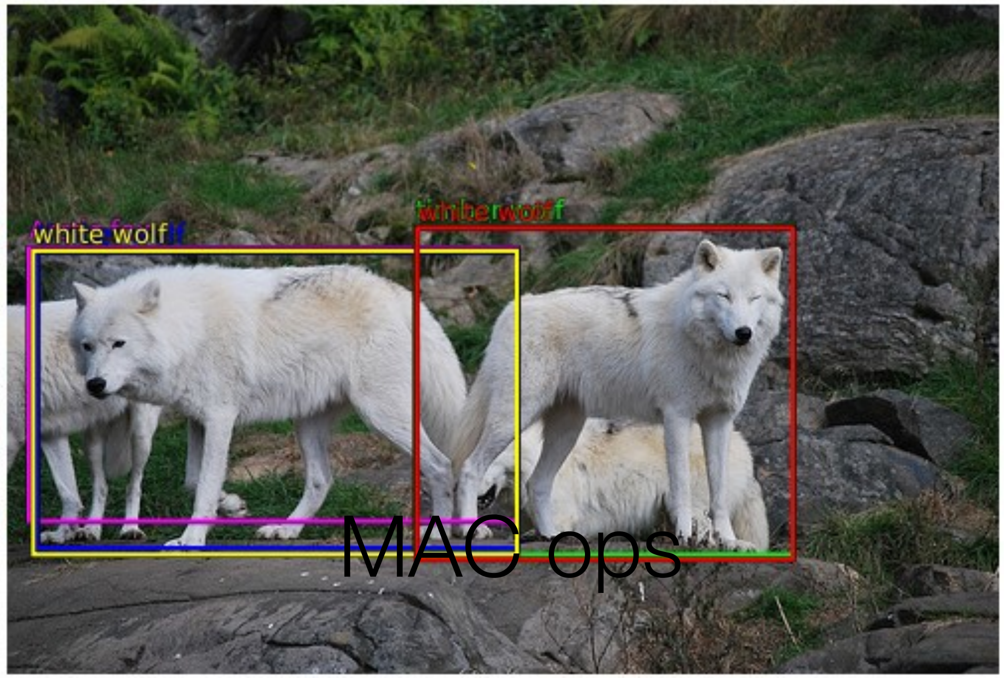
Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Credit: Y. LeCun, Facebook

Applications

Object Recognition in Images

Object Recognition



Top 5:
white wolf
white wolf
timber wolf
timber wolf
Arctic fox

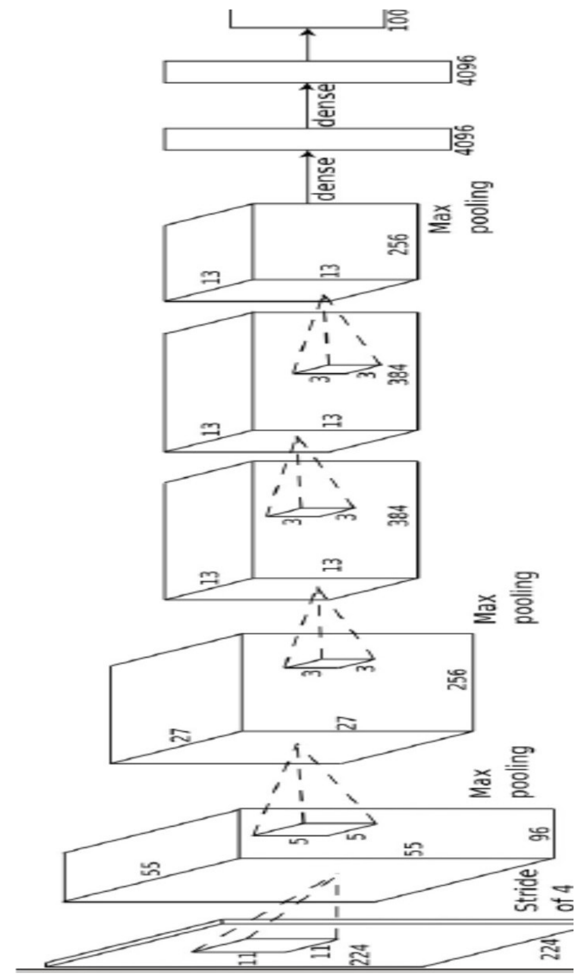
ILSVRC2012_val_00000027.JPEG

Credit: Y. LeCun, Facebook

Object Recognition [Krizhevsky, Sutskever, Hinton 2012]

– Won the 2012 ImageNet LSVRC. 60 Million parameters, 832M Matrix multiplication + accumulation operations

4M	FULL CONNECT	4Mflop
16M	FULL 4096/ReLU	16M
37M	FULL 4096/ReLU	37M
	MAX POOLING	
442K	CONV 3x3/ReLU 256fm	74M
1.3M	CONV 3x3ReLU 384fm	224M
884K	CONV 3x3/ReLU 384fm	149M
	MAX POOLING 2x2sub	
	LOCAL CONTRAST NORM	
307K	CONV 11x11/ReLU 256fm	223M
	MAX POOL 2x2sub	
	LOCAL CONTRAST NORM	
35K	CONV 11x11/ReLU 96fm	105M



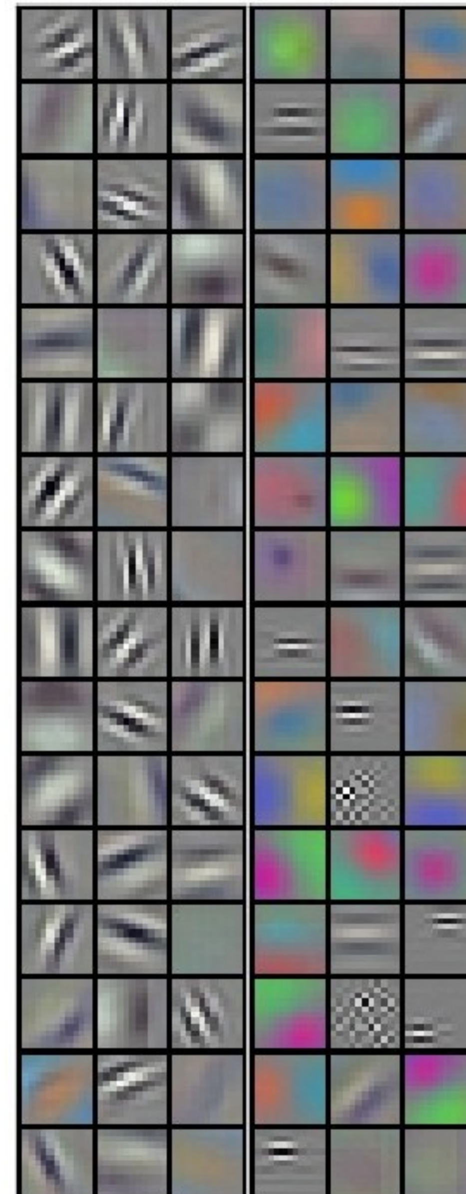
Credit: Y. LeCun, Facebook

Object Recognition [Krizhevsky, Sutskever, Hinton 2012]

- Method: large convolutional net
  - 650K neurons, 832M synapses, 60M parameters
  - Trained with backprop on NVIDIA GPU
  - Trained “with all the tricks Yann came up with in the last 20 years, plus dropout” (Hinton, NIPS 2012)
  - Rectification, contrast normalization,...
- Error rate: 15% (whenever correct class isn't in top 5) Previous state of the art: 25% error

- **A revolution in computer vision**

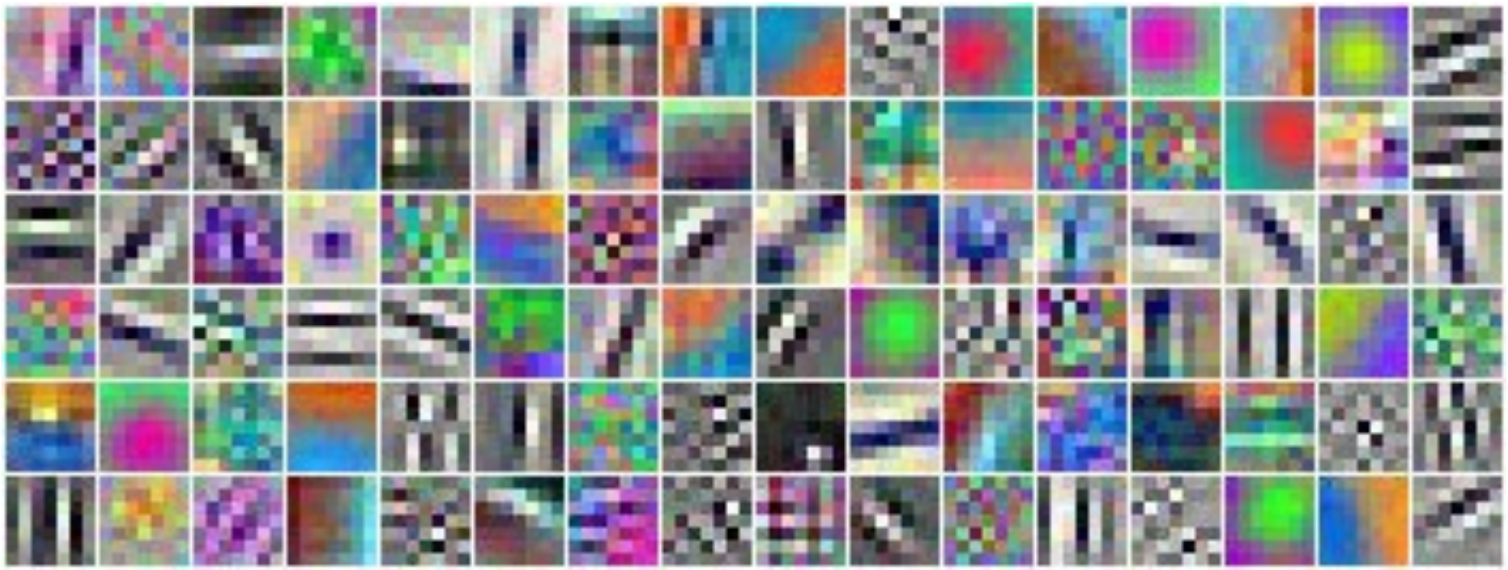
- Acquired by Google in Jan 2013
- Deployed in Google+ Photo Tagging in May 2013



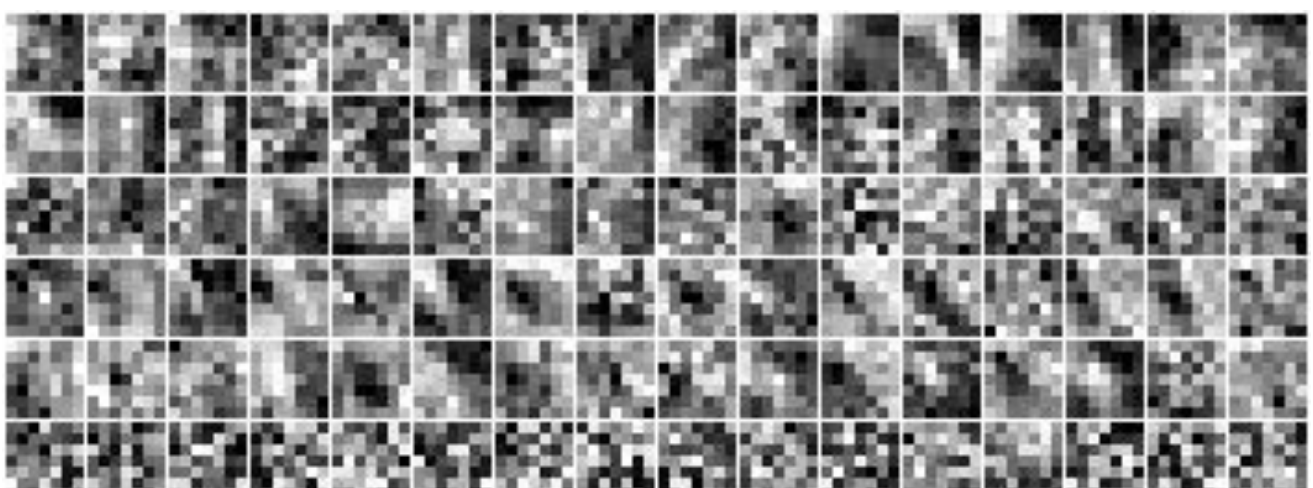
Credit: Y. LeCun, Facebook

Filters: Layer 1 (7x7) and Layer 2 (7x7)

– Layer 1: 3x96 filters, RGB->96 feature maps, 7x7 Filters, stride 2



– Layer 2: 96x256 filters, 7x7

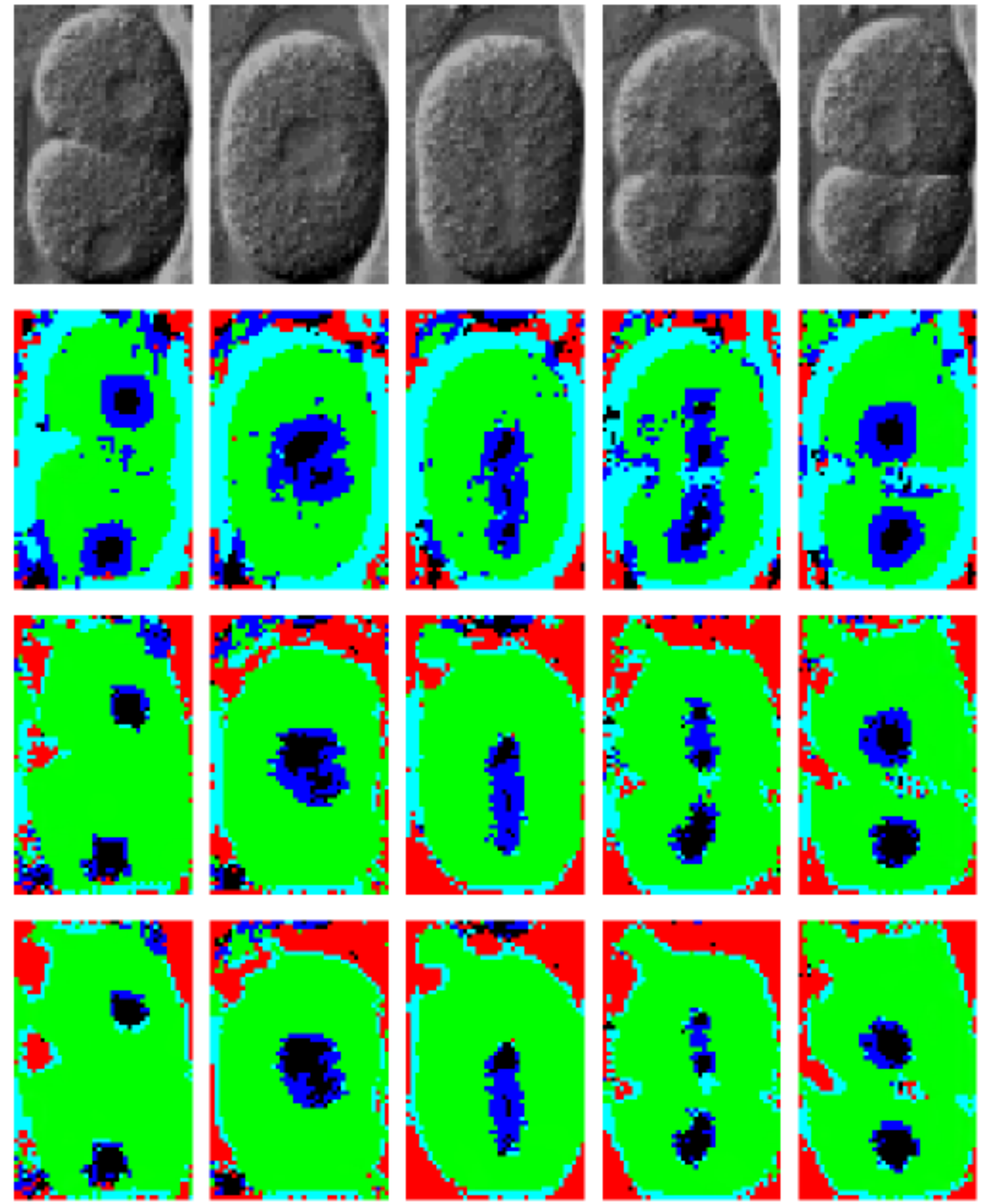


Credit: Y. LeCun, Facebook

Image Segmentation

ConvNets for image segmentation

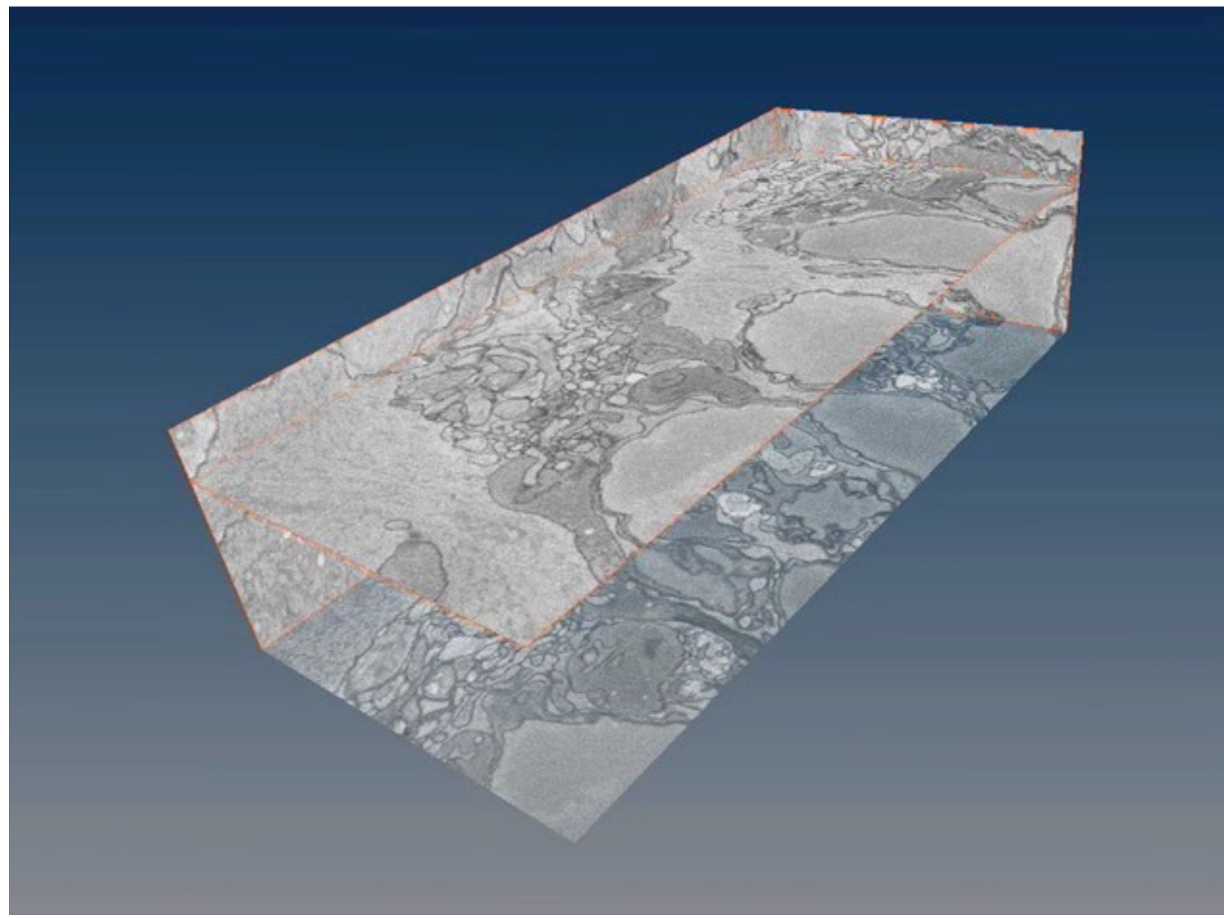
- Biological Image Segmentation
  - [Ning et al. IEEE-TIP 2005]
- Pixel labeling with large context using a convnet
- ConvNet takes a window of pixels and produces a label for the central pixel



Credit: Y. LeCun, Facebook

ConvNet in Connectomics [Jain, Turaga, Seung 2007-present]

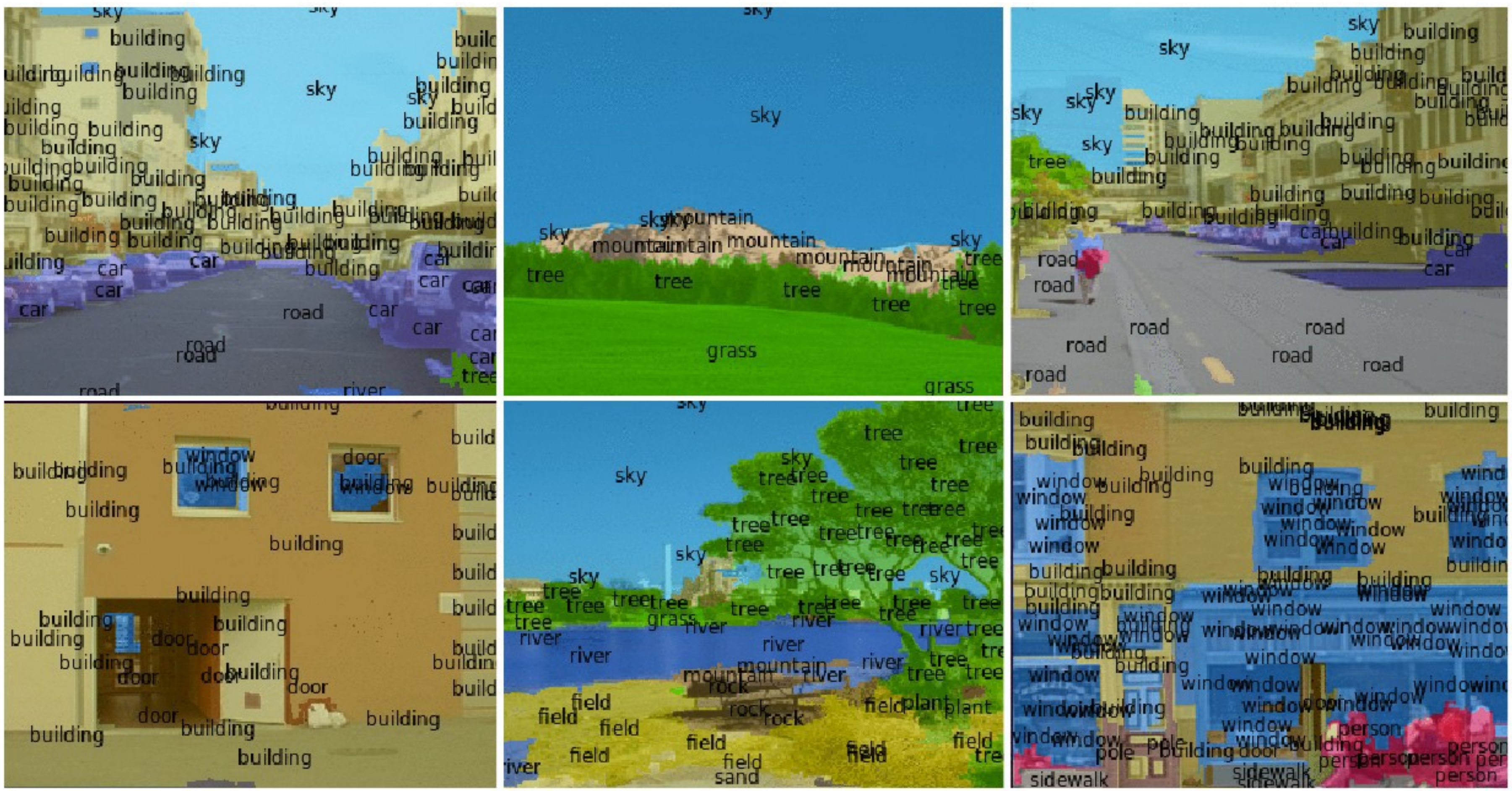
– 3D ConvNet Volumetric Images
– Each voxel labeled as "membrane" or "non-membrane using a 7x7x7 voxel neighborhood
– Has become a standard method in connectomics



Semantic labeling / scene parsing:

Labeling every pixel with the object it belongs to

– Would help identify obstacles, targets, landing sites, dangerous areas
  Would help line up depth map with edge maps



[Farabet et al. ICML 2012, PAMI 2013]

Credit: Y. LeCun, Facebook