Computer Vision

Neurobio 230
Bill Lotter

Exciting time: Neuroscience ⇔ computer vision

-Traditionally: computer vision relied on hand crafted features

-Today: “Deep Learning”

-loosely based on how the brain does computations

-most of components learned from data

-a lot of commonalities between computer vision models and the visual ventral stream in the brain

Overview of Computer Vision Problems

Object Recognition

Image Segmentation

Optical Character Recognition

Face Identification

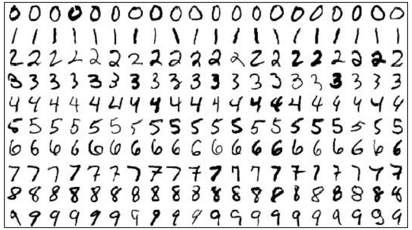
Action Recognition

…

Applications to: photography, self-driving cars, medical imaging analysis,..

Common Testbeds for Computer Vision

MNIST



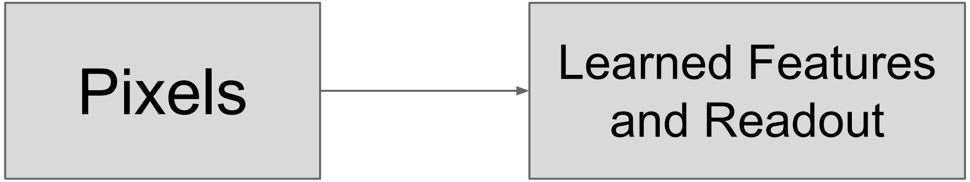
LFW

Imagenet

General Problem Formulation

Pre ~2012:

Post 2012:



Focusing on Object Recognition: Convolutional Neural Networks (CNNs)

Background:

Hubel and Wiesel Simple and Complex Cells (1959, 1960s)

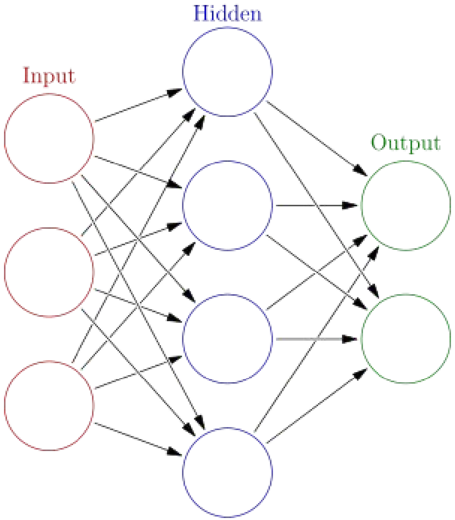
Neocognitron (Fukushima, 1980)

HMAX (Riesenhuber & Poggio 1999, Serre, Kreiman et al. 2007)

Yann LeCun’s work on MNIST with CNNs (1998)

What is an Artificial Neural Network?

a lot of variations, hard to generalize, but a simple ANN looks something like this..

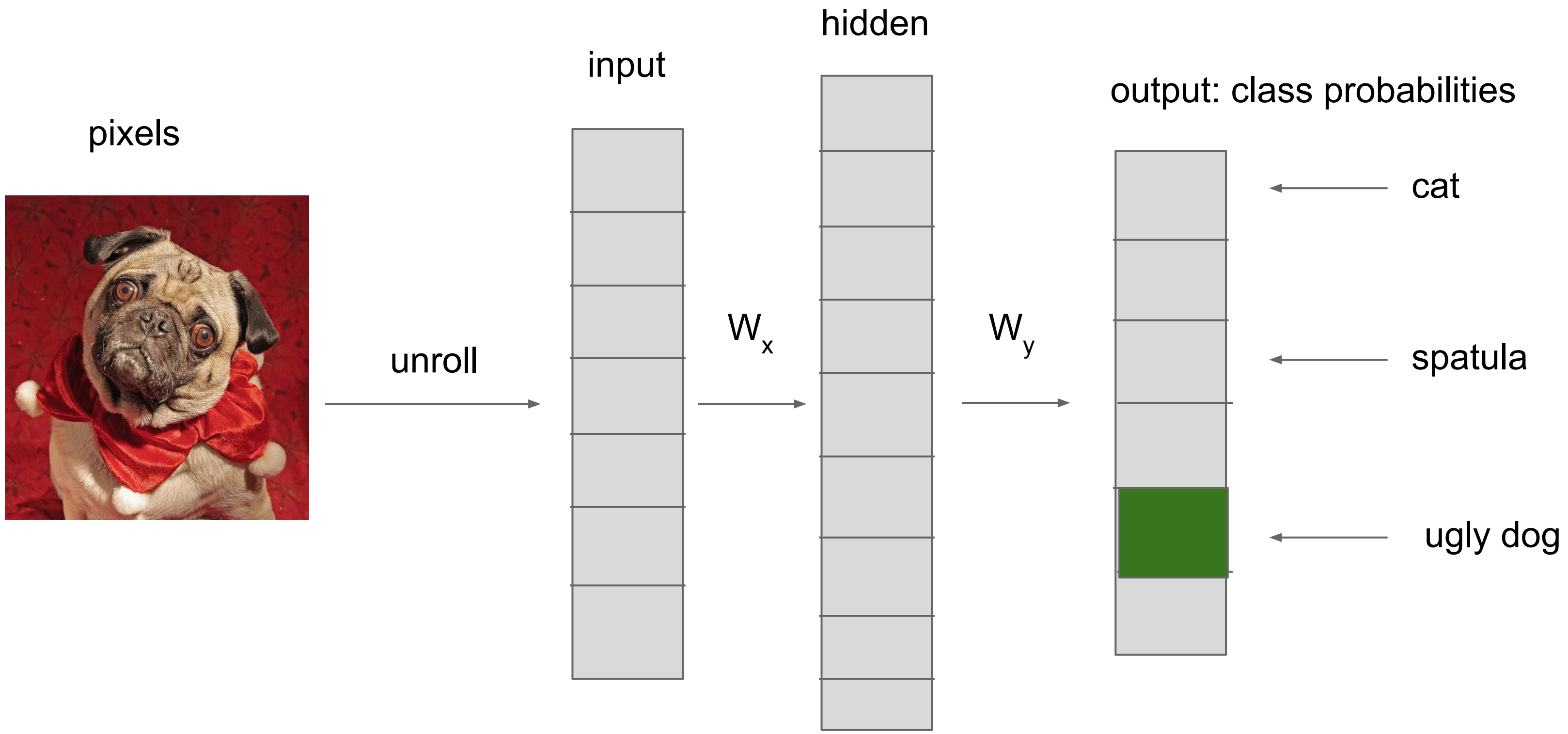


Training the Network: Backprop

Backpropagation (Rumelhart, Hinton, Williams 1986): way to calculate gradient of error in terms of network parameters

Today: gradient descent with some bells and whistles

Formulating for object recognition...



Taking a look at parameters..

image: 256x256x3 = 196,608 inputs

outputs: 1000 categories

even if just go directly from image to outputs:

1000 x 196,608 = 196 million params!!

even if you have 1 million training images, you would severely overfit the network

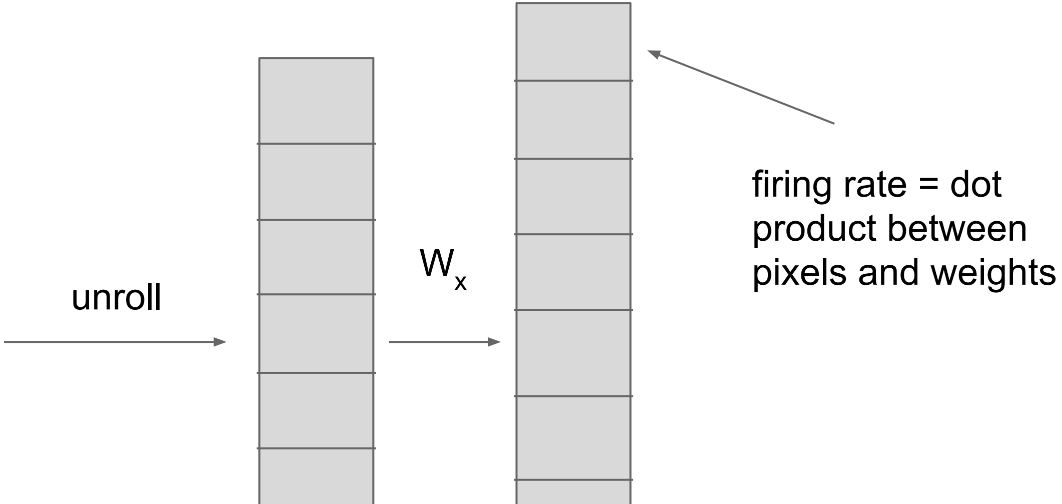
Using Convolutions

Natural images aren’t just random arrays, they have structure

Two things to exploit while designing networks: locality and ~spatial invariance

Relating to neuroscience: weights for a given unit can be thought of as receptive field

pixels

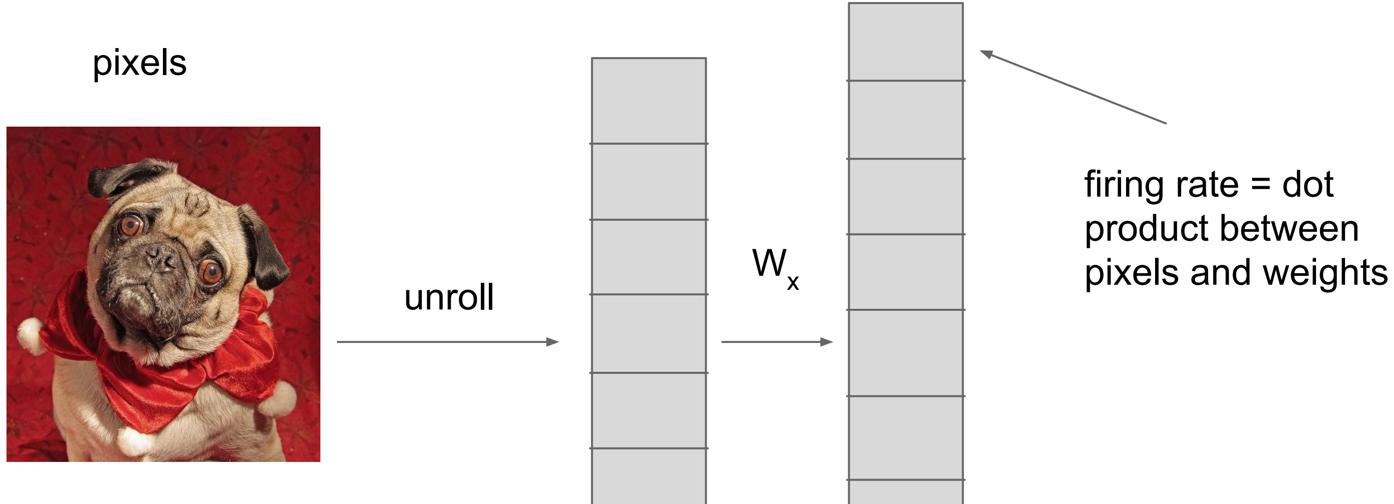


Using Convolutions

Natural images aren't just random arrays, they have structure

Two things to exploit while designing networks: locality and ~spatial invariance

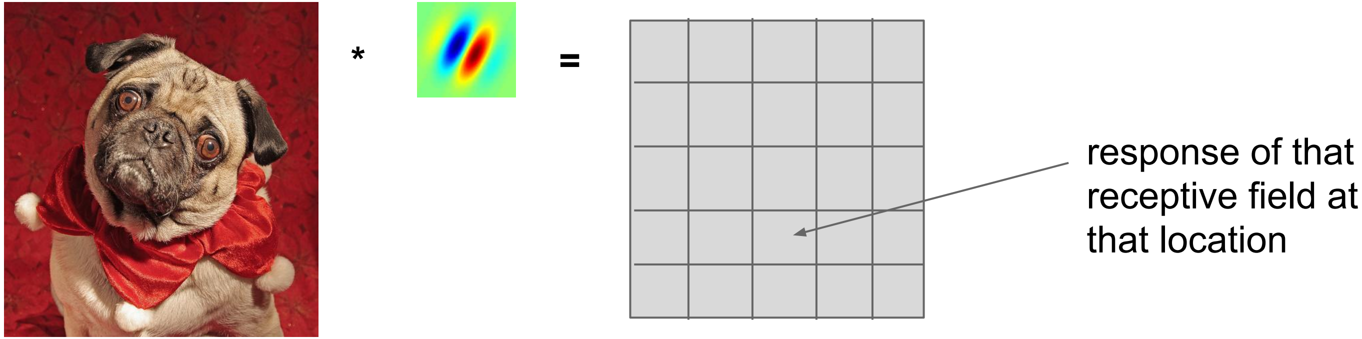
Relating to neuroscience: weights for a given unit can be thought of as receptive field



Using Convolutions

Weights as receptive fields: localized and can replicate over visual field

=> It makes sense to use convolutions

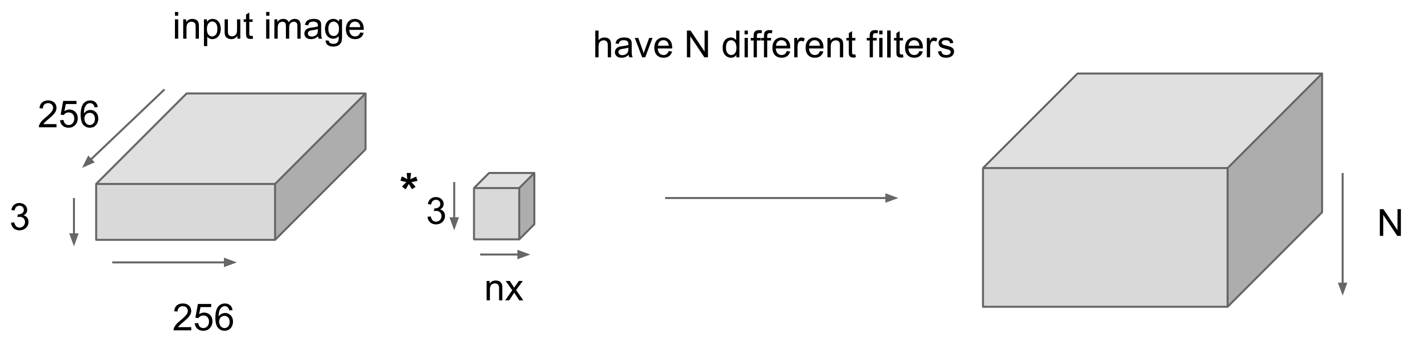


Using Convolutions

Full formulation: layers have “depth” as well

(x, y) pixel position and 3 color channels

We want a bunch of different filters to convolve the image with



Incorporating other stuff we know is important in biology

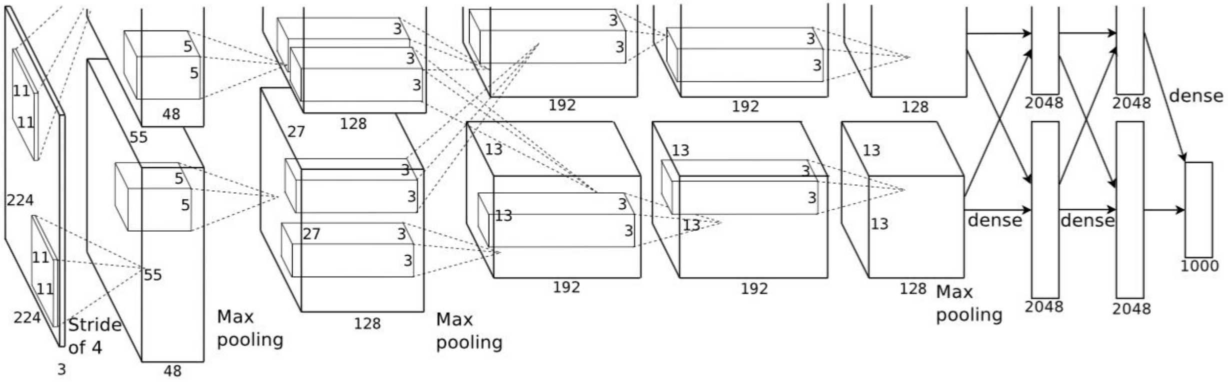
Hierarchy: ventral stream has several layers (V1, V2,...)

Neurons are non-linear: common non-linearity used today is rectified linear units (don't allow neurons to have negative firing rate)

"Complex"-type cells: incorporating pooling

Putting it all together…

Krizhevsky et al. 2012 (Alexnet)



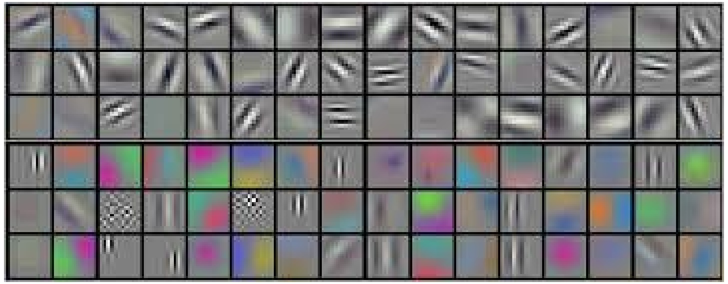
Comparing with Biology

Similarities

hierarchical

receptive fields get bigger as go higher

first layer trained weights look like V1 receptive fields



Differences

backprop

supervised vs. unsupervised learning

final model is purely feedforward

Other Cool Stuff

Learned feature representations are generalizable

can do other tasks like object localization (Oquab et al. 2015)

people use Alexnet feature representations as input to many other problems

Inverting convolutional neural networks

train another network to go from feature

representation back to pixel space (Dosovitskiy 2015)

can see what different layers represent

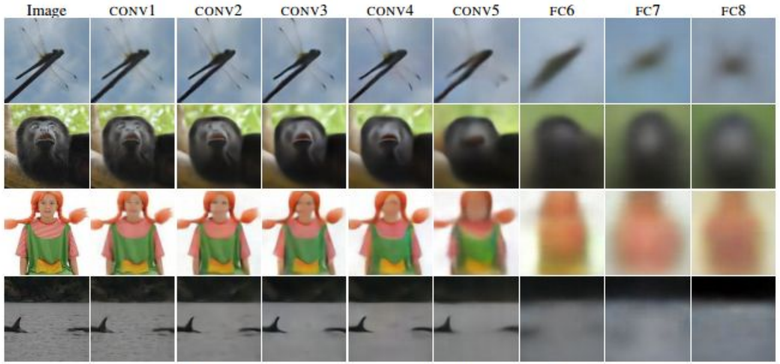
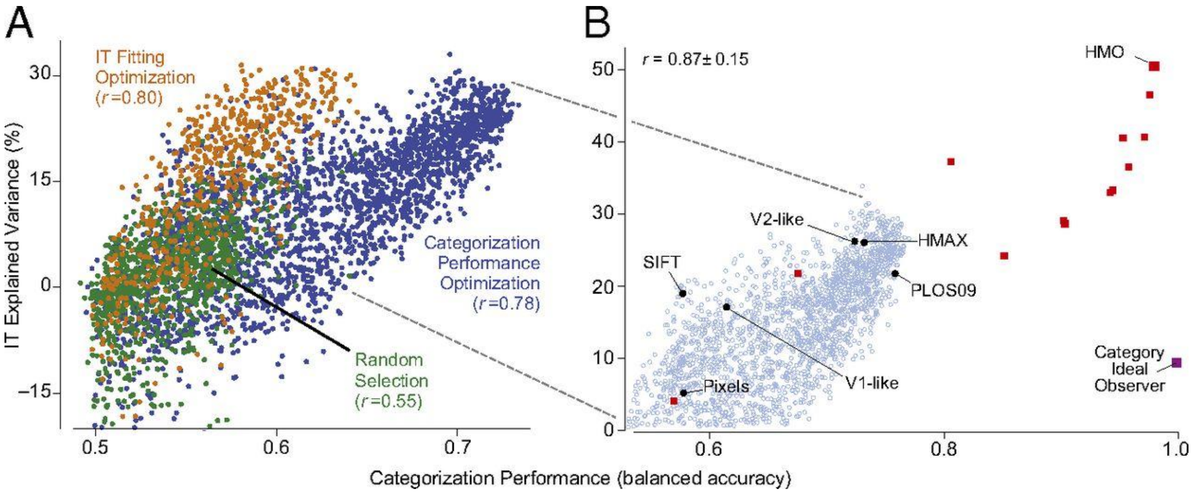


Figure 5: Reconstructions from different layers of AlexNet.

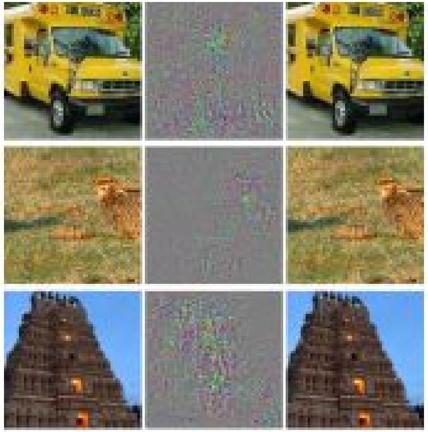
Other Cool Stuff

The more predictive a model is of neural data, the better it is at performance (Yamins 2014)



Other Cool Stuff

Nonetheless, it is easy to fool convnets (Szegedy 2013)



(a)

classified as ostrich

Final Thoughts

Still far away from making machines that can perform as well as humans, but making steady progress by designing models that share many features with brain

Neuroscience has informed computer vision, but computer vision models also allow for testing of neuroscience theories

much easier to do “neuroscience” on models than real brains