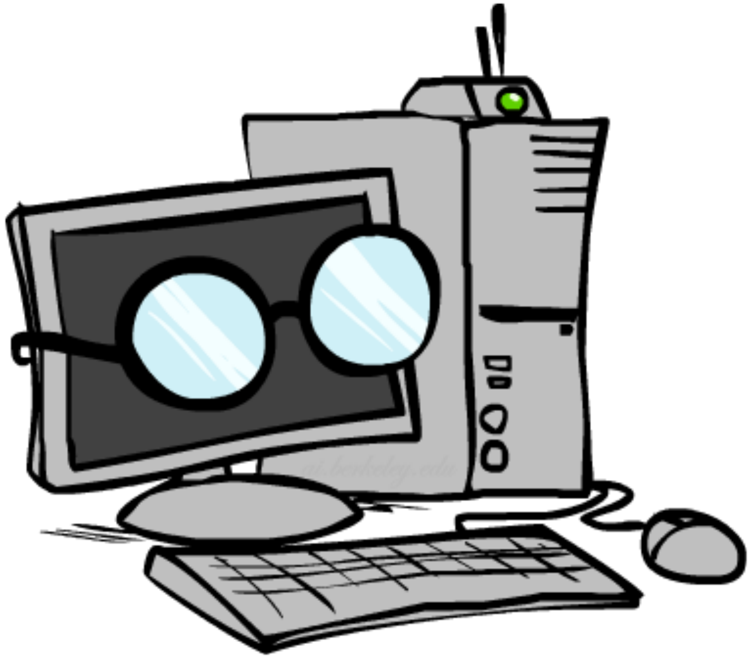
Machine Learning

Convolutional Neural Networks

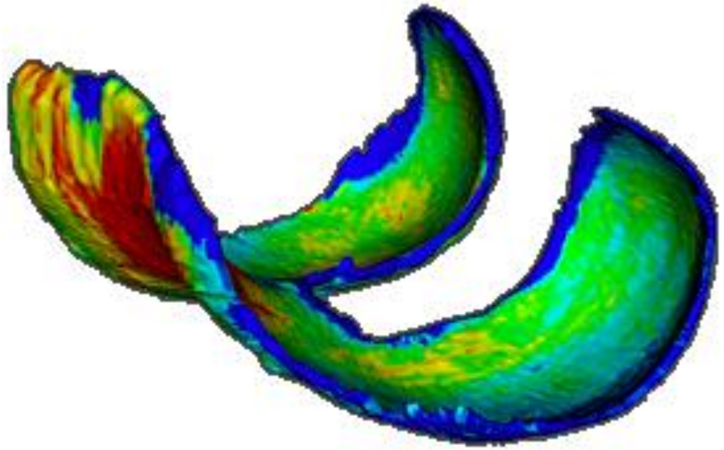


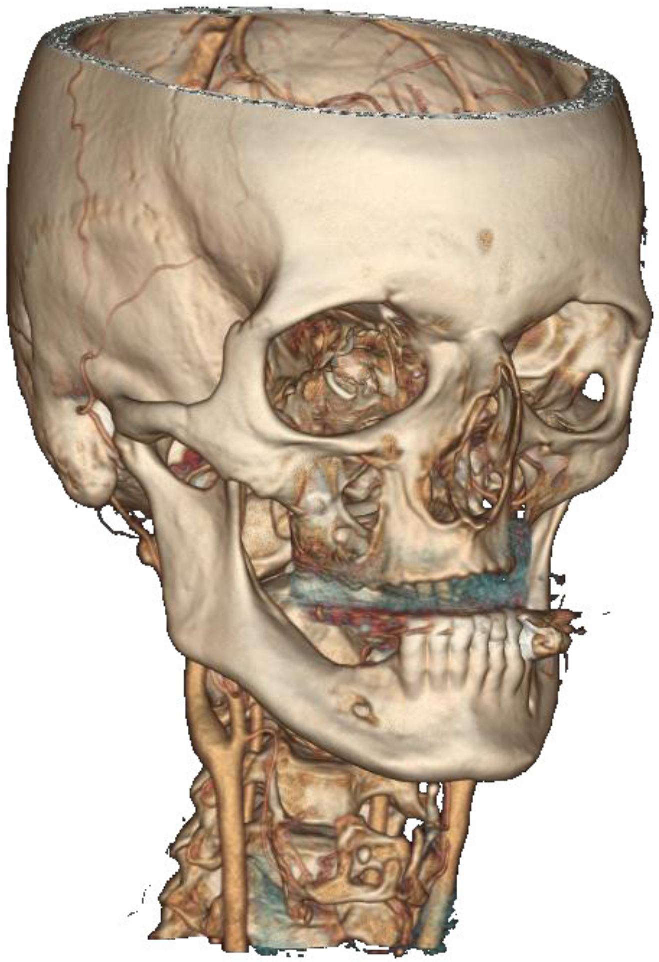
Pat Virtue

University of California, Berkeley

Fun drawing credits: http://ai.berkeley.edu & http://csillustrated.berkeley.edu

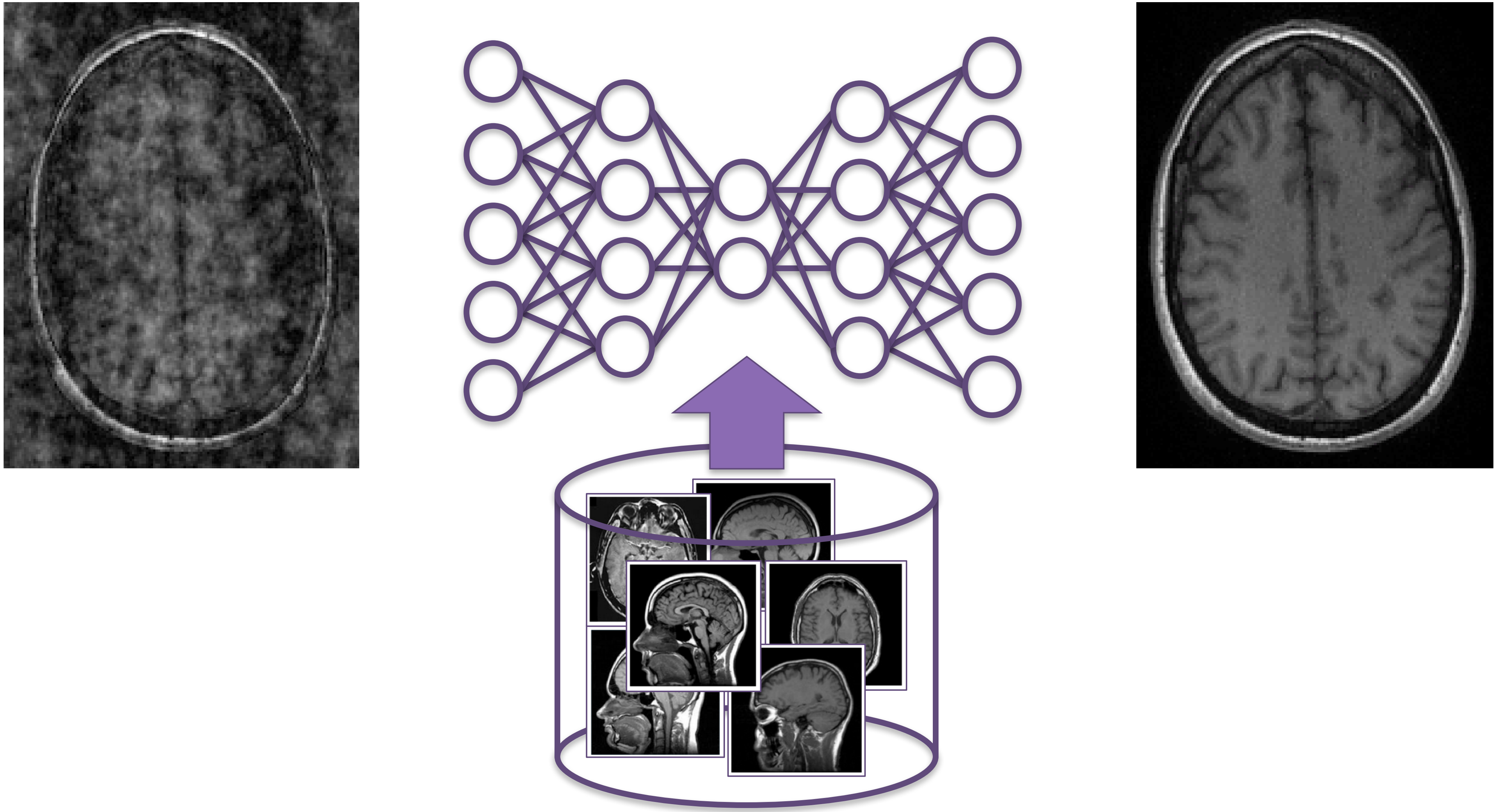
“Allow myself to introduce… myself” – *A. Powers*



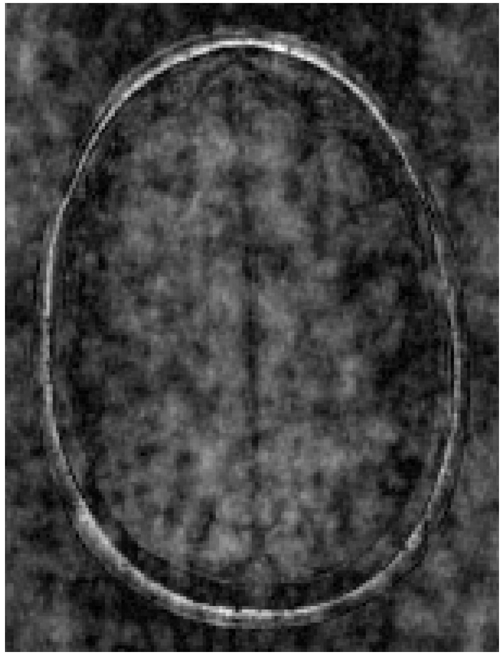


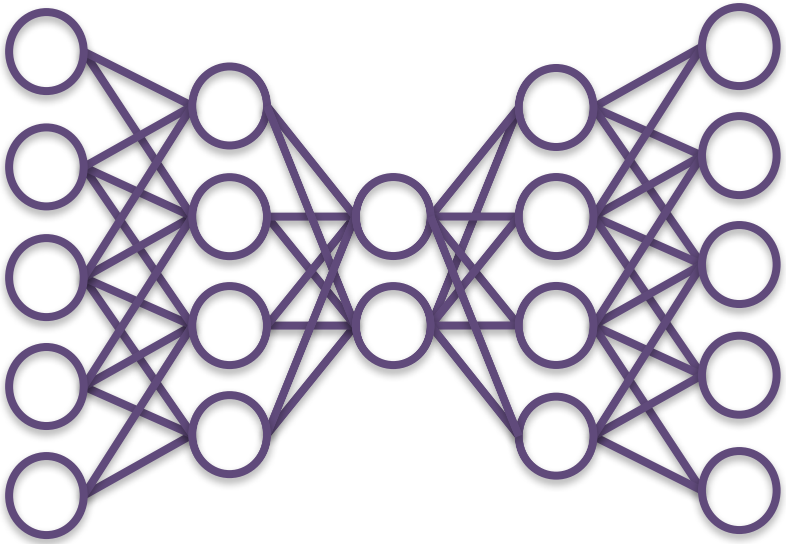
Images: GE Healthcare

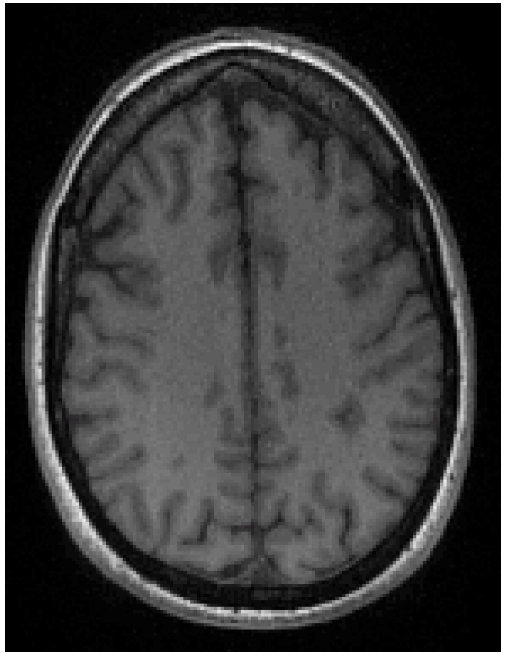
“Allow myself to introduce… myself” – *A. Powers*



"Allow myself to introduce… myself" – *A. Powers*





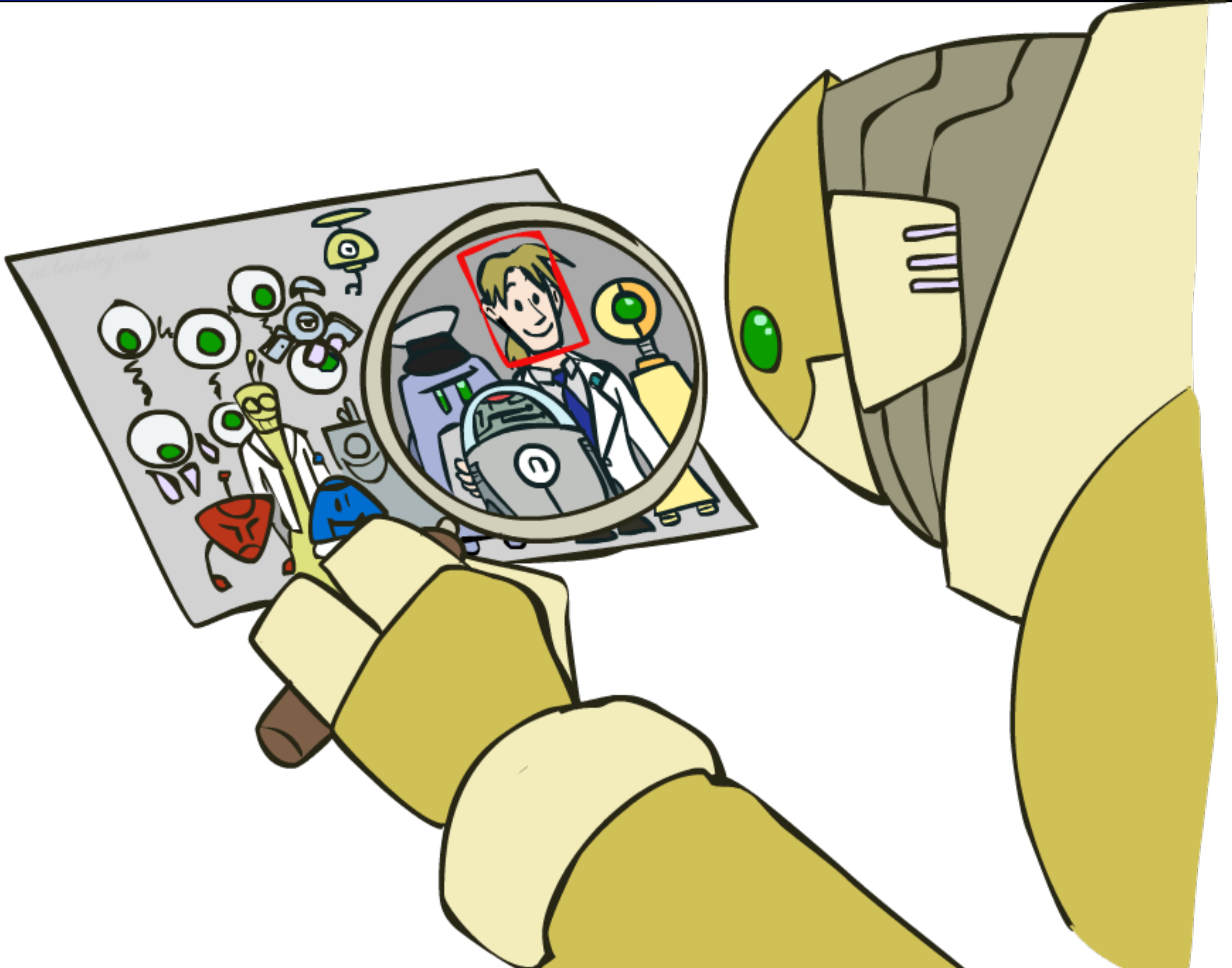


UC Berkeley CS188 Intro to AI

Outline

1. Measuring the current state of computer vision
2. Why convolutional neural networks
   - Old school computer vision
   - Image features and classification
3. Convolution “nuts and bolts”

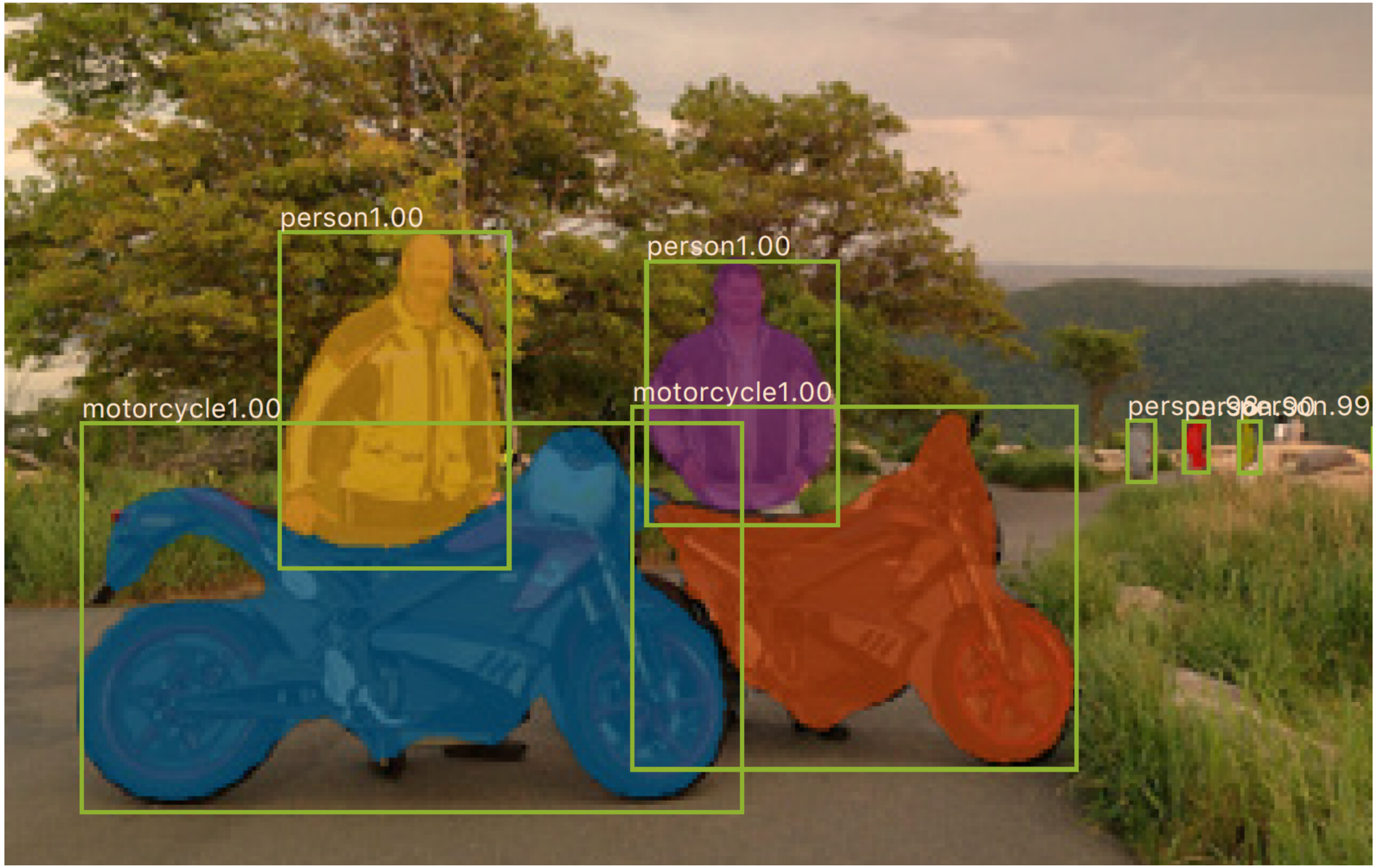
Computer Vision: How far along are we?



Computer Vision: How far along are we?

Terminator 2, 1991

Computer Vision: How far along are we?



0.2 seconds per image

Mask R-CNN

He, Kaiming, et al. "Mask R-CNN." *Computer Vision (ICCV), 2017 IEEE International Conference on*. IEEE, 2017.

Computer Vision: How far along are we?

“My CPU is a neural net processor, a learning computer”

Terminator 2, 1991

Computer Vision: Autonomous Driving

Tesla, Inc: https://vimeo.com/192179726

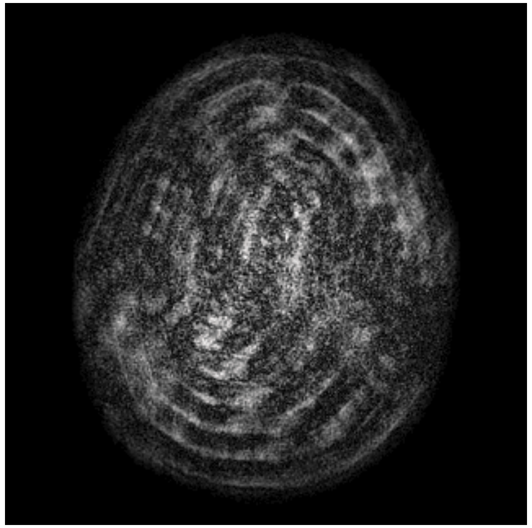
Computer Vision: Domain Transfer

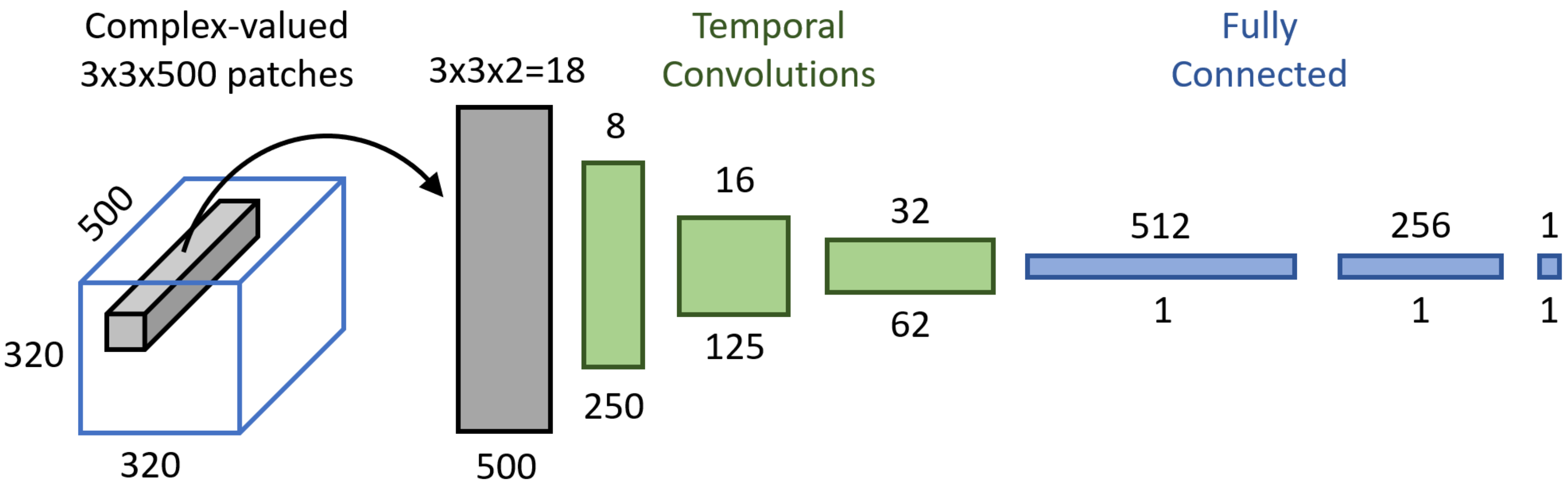
CycleGAN

Jun-Yan Zhu*, Taesung Park*, Phillip Isola, and Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", ICCV 2017.

Temporal Convolution

MR Fingerprinting





Patrick Virtue , Jonathan I Tamir , Mariya Doneva , Stella X Yu , and Michael Lustig. "Learning Contrast Synthesis from MR Fingerprinting", ISMRM 2018, forthcoming.

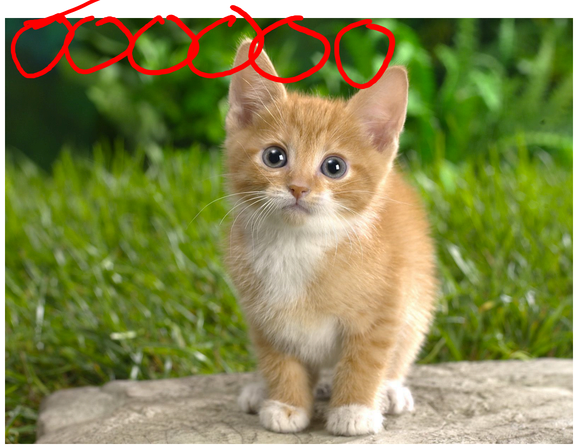
Outline

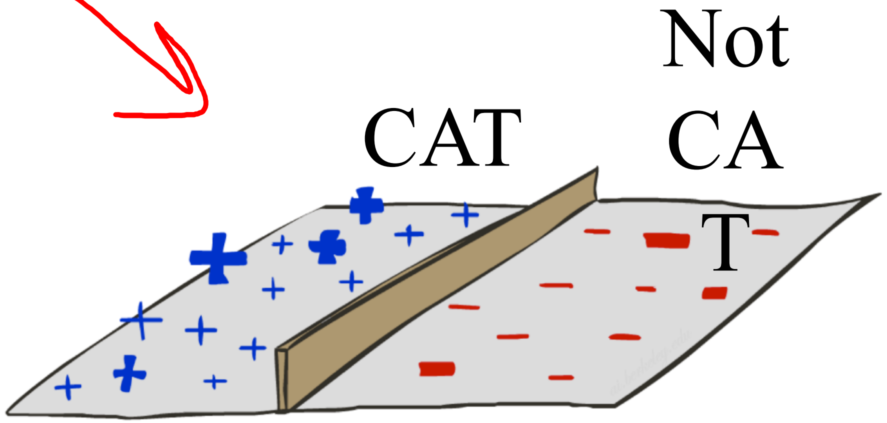
1. Measuring the current state of computer vision
2. Why convolutional neural networks
   - Old school computer vision
   - Image features and classification
3. Convolution “nuts and bolts”

Image Classification

- What's the problem with just directly classifying raw pixels in high dimensional space?

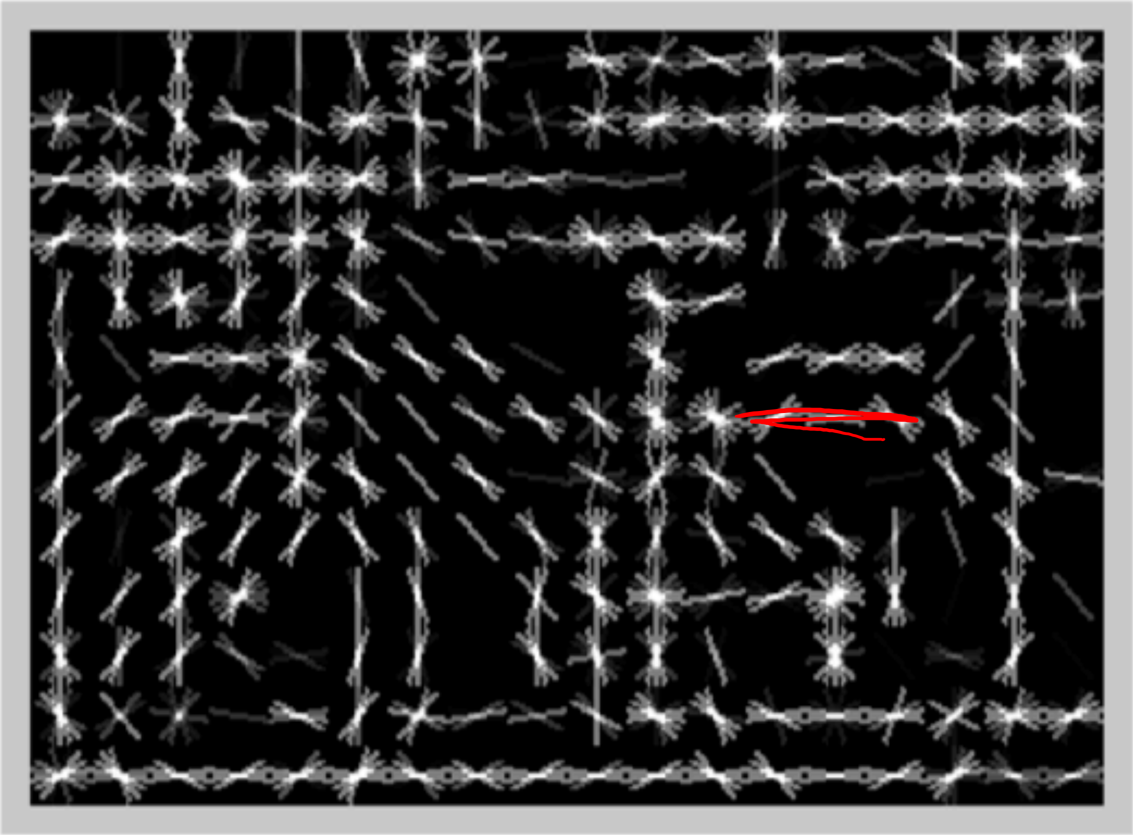
256^2 - D





CAT

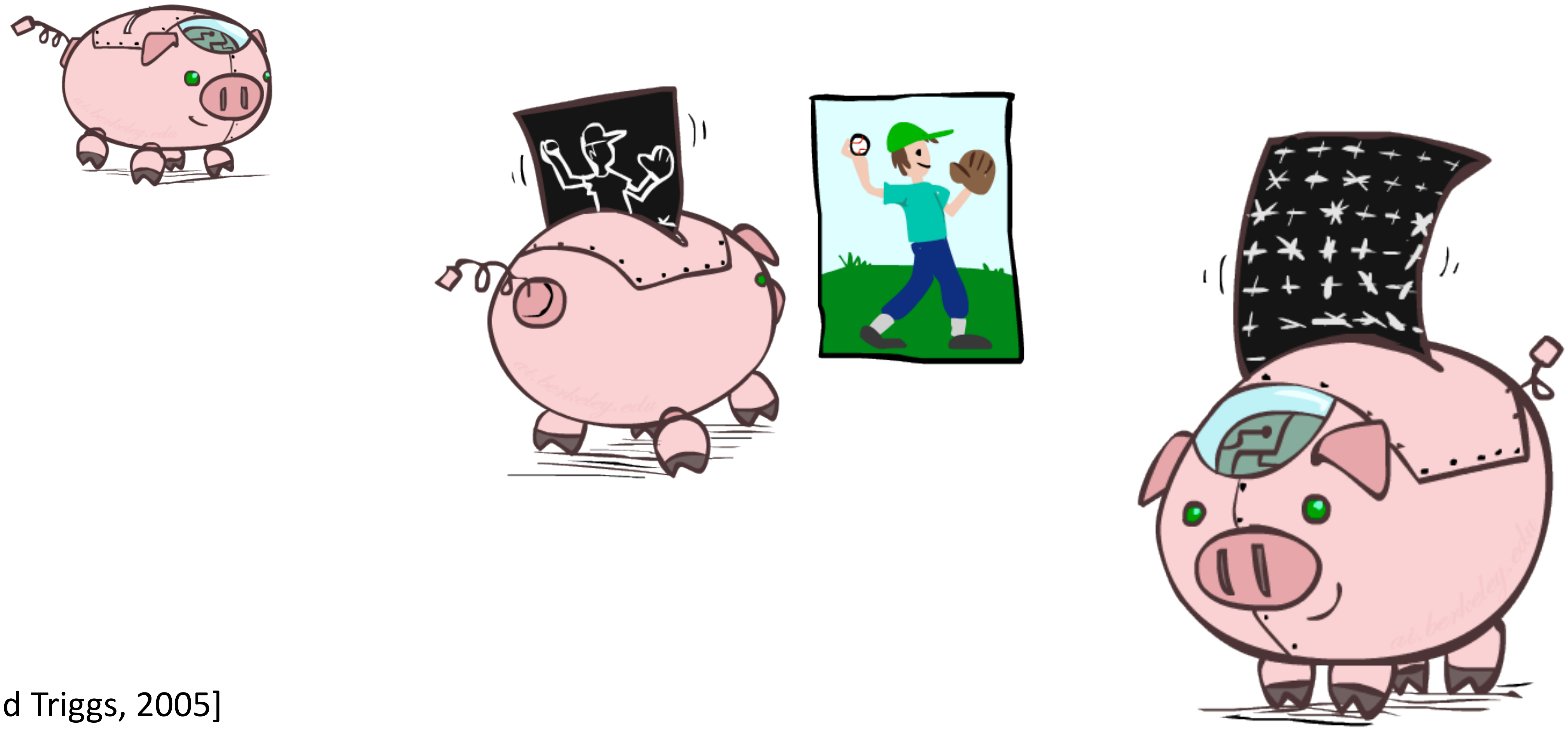
Image Classification



[Dalal and Triggs, 2005]

HoG Filter

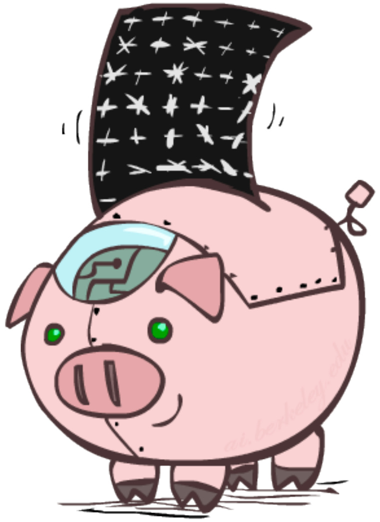
- HoG: Histogram of oriented gradients

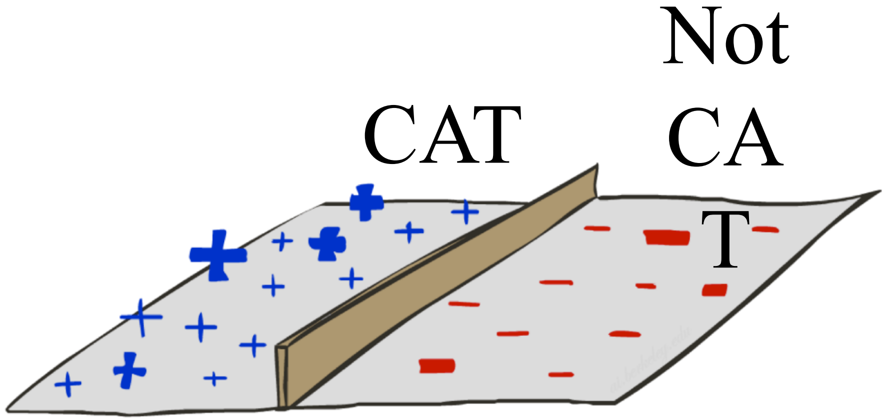


[Dalal and Triggs, 2005]

Image Classification

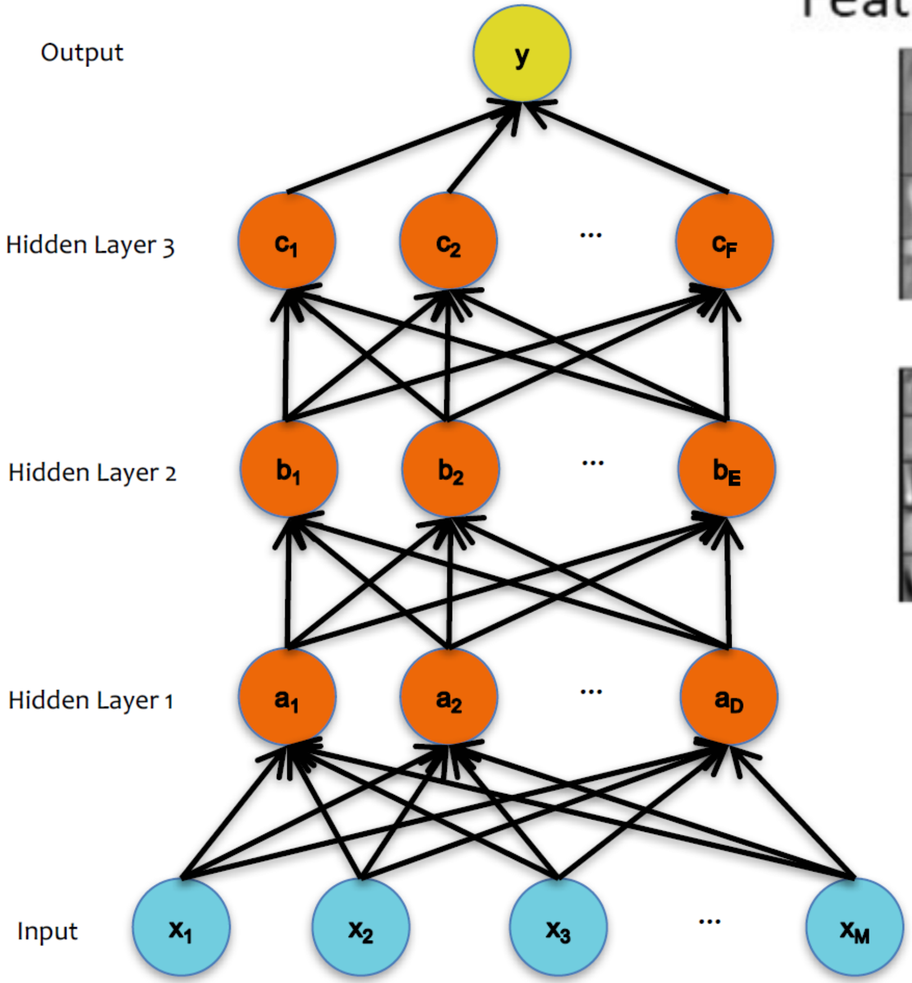
- HOG features passed to a linear classifier (SVM)



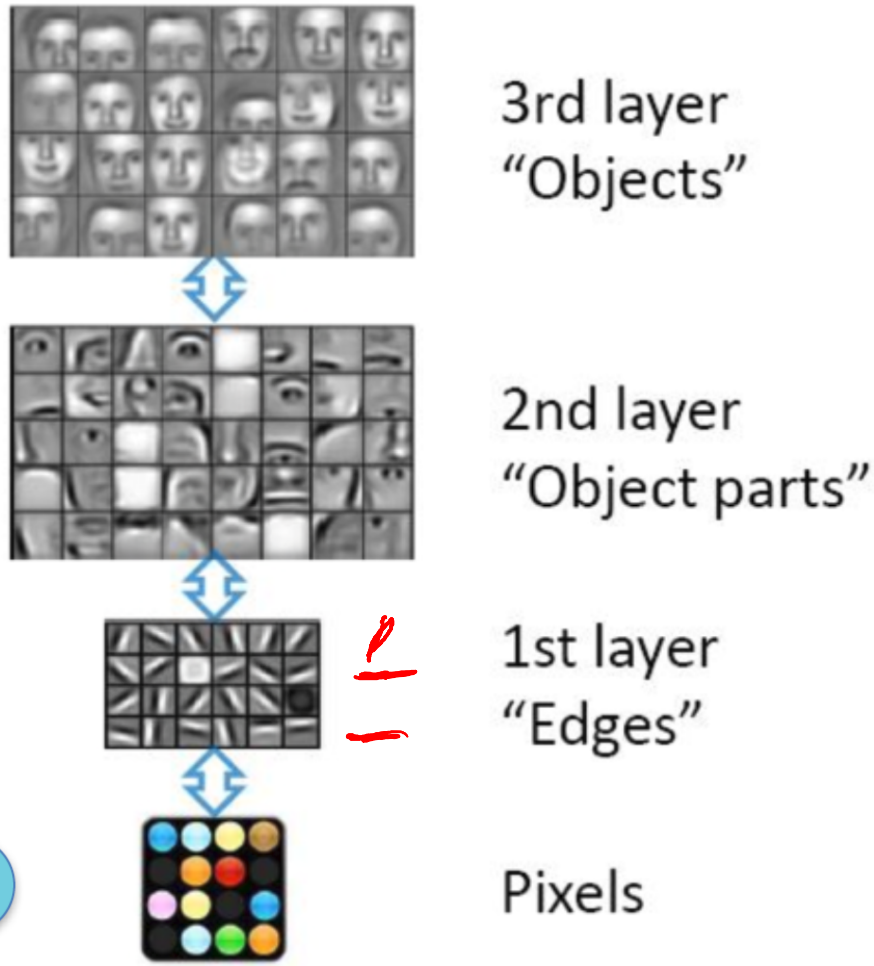


CAT

Classification: Learning Features

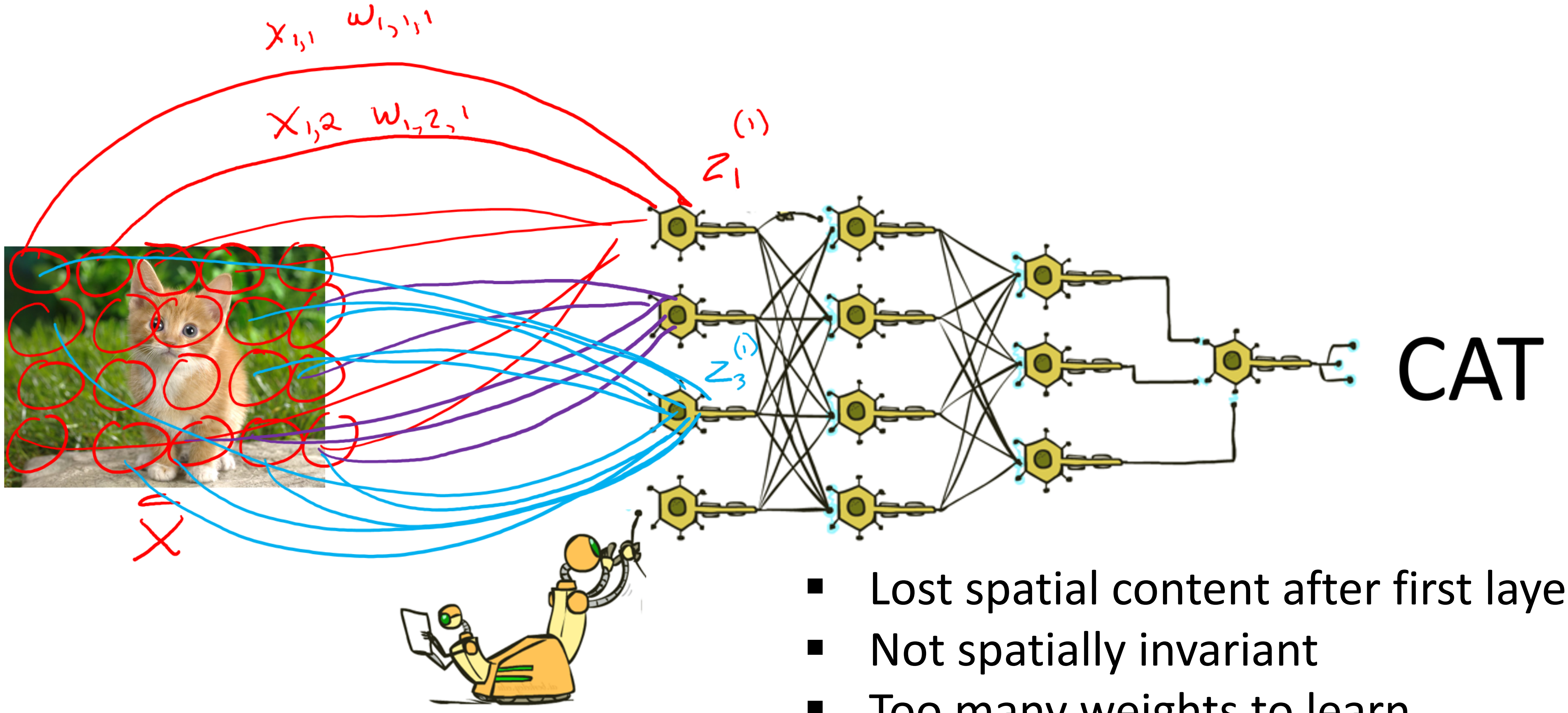


Feature representation



Example from Honglak Lee (NIPS 2010)

Classification: Deep Learning



- Lost spatial content after first layer
- Not spatially invariant
- Too many weights to learn

Convolution

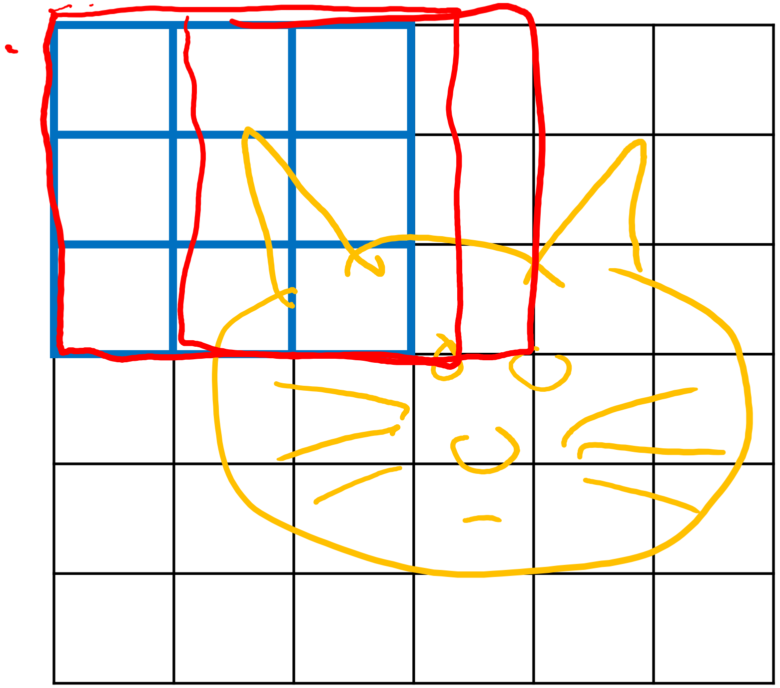
- Signal processing definition

$$z[i,j] = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} x[i-u, j-v] \cdot w[u, v]$$

-1	0	1
-2	0	2
-1	0	1

- Relaxed definition
  - Drop infinity; don't flip kernel

$$z[i,j] = \sum_{u=0}^{K-1} \sum_{v=0}^{K-1} x[i+u, j+v] \cdot w[u, v]$$



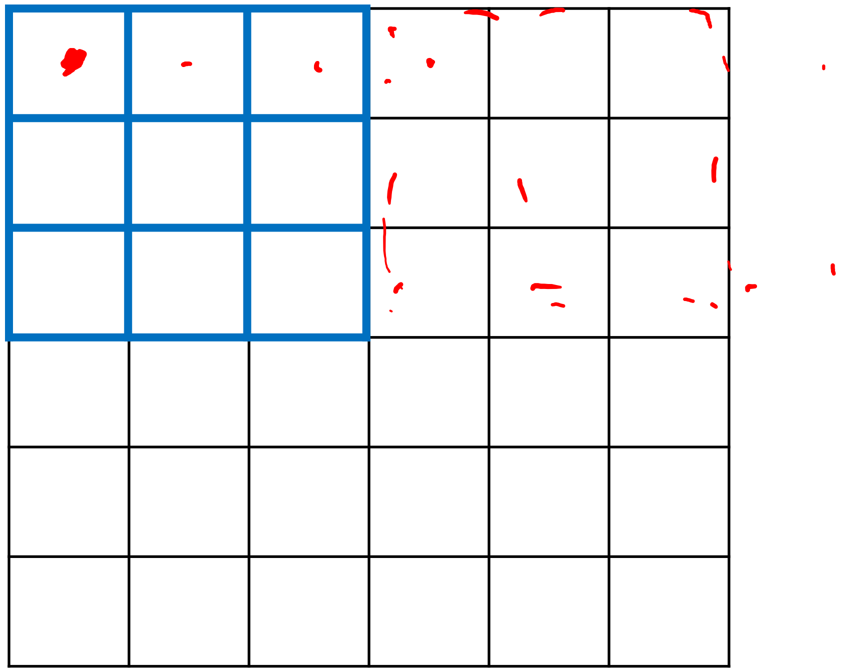
Convolution

- Relaxed definition

$$z[i,j] = \sum_{u=0}^{K-1} \sum_{v=0}^{K-1} x[i+u, j+v] \cdot w[u\,, v]$$

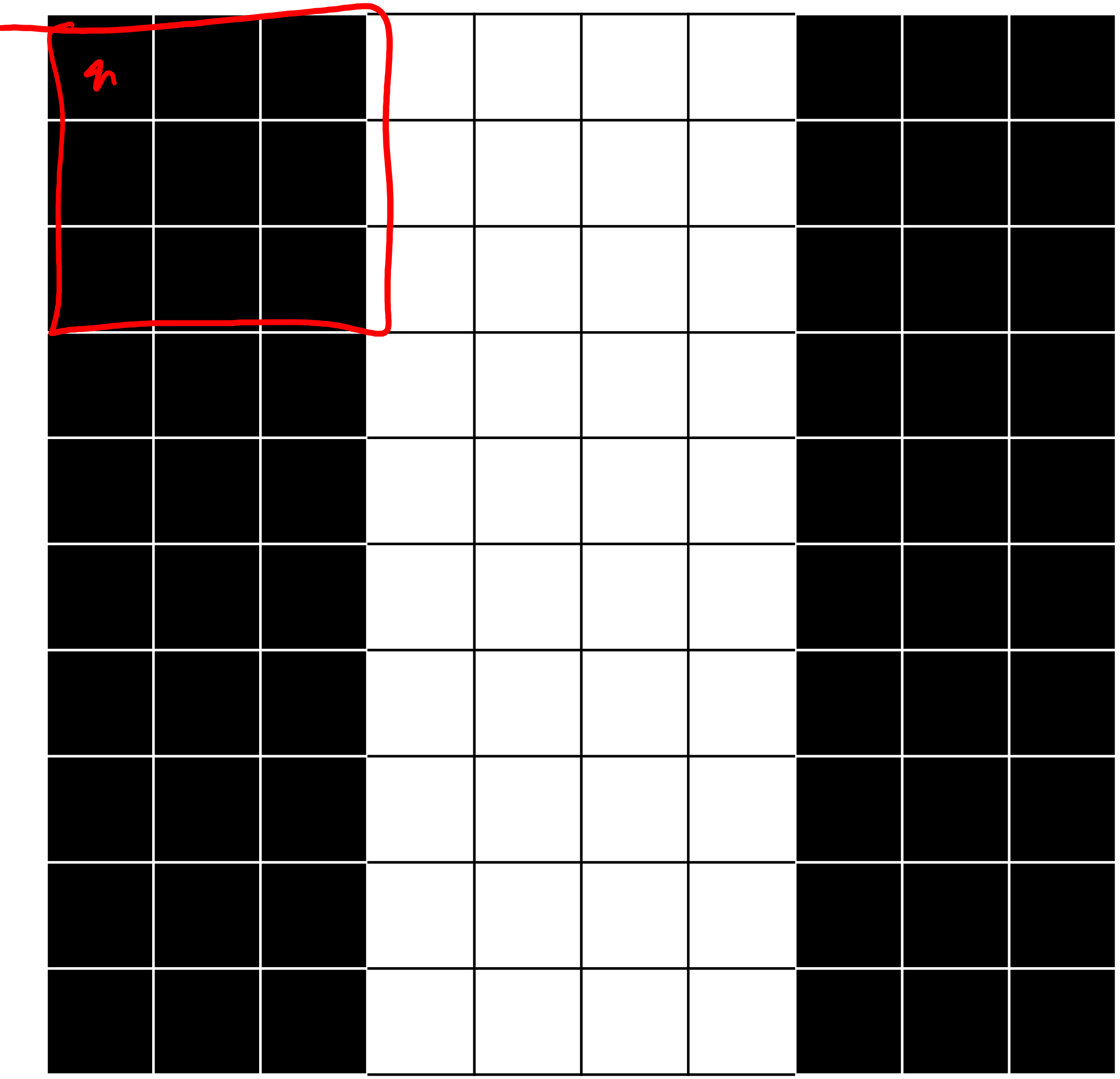
-1	0	1
-2	0	2
-1	0	1

```
for i in range(0, im_width - K + 1):
  for j in range(0, im_height - K):
    im_out[i,j] = 0
    for u in range(0, K):
      for v in range(0, K):
        im_out[i,j] += im[i+u, j+v] * kernel[u,v]
```



GPU!!

Convolution



-1	0	1
-1	0	1
-1	0	1

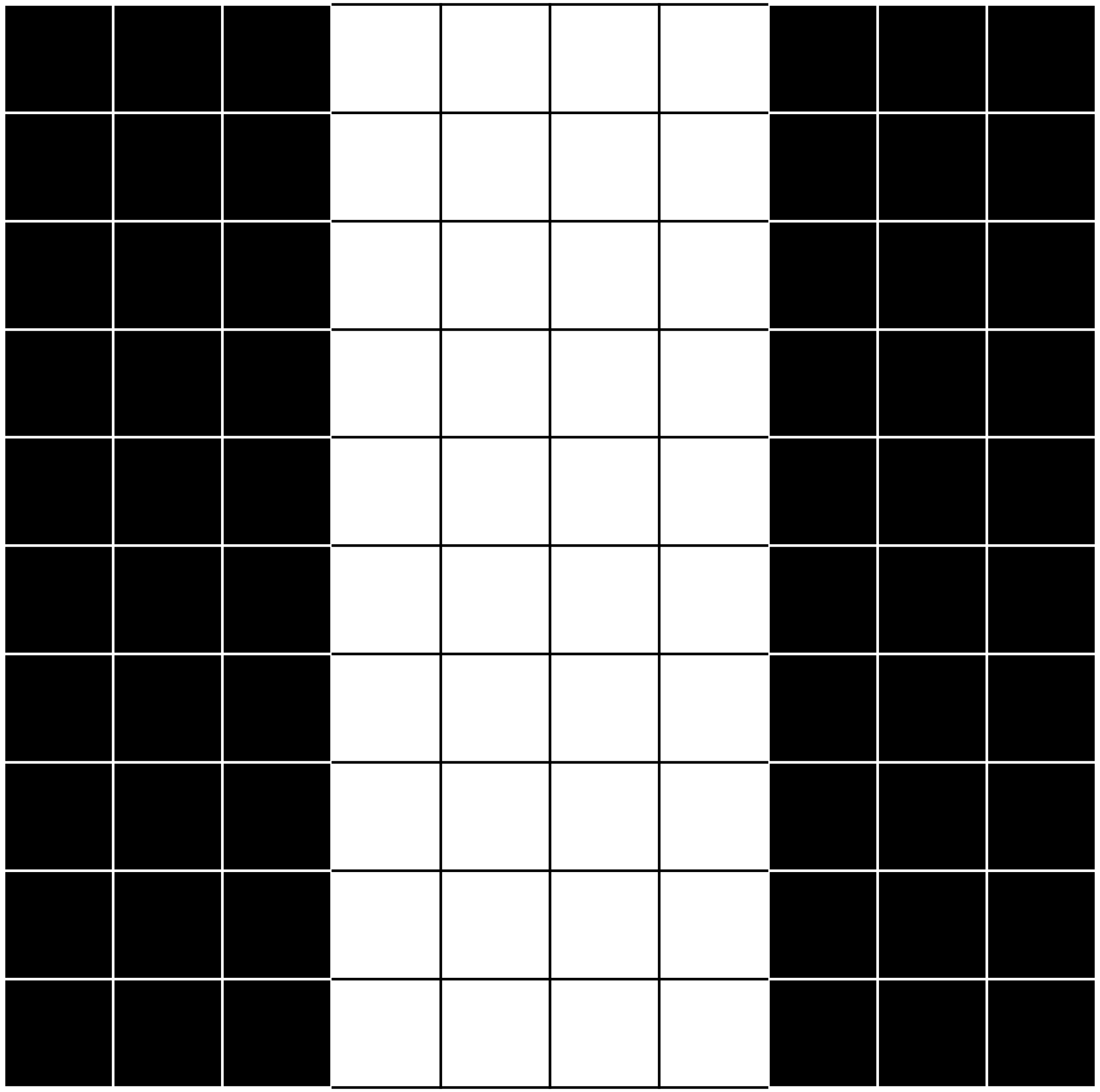
Convolution

0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0

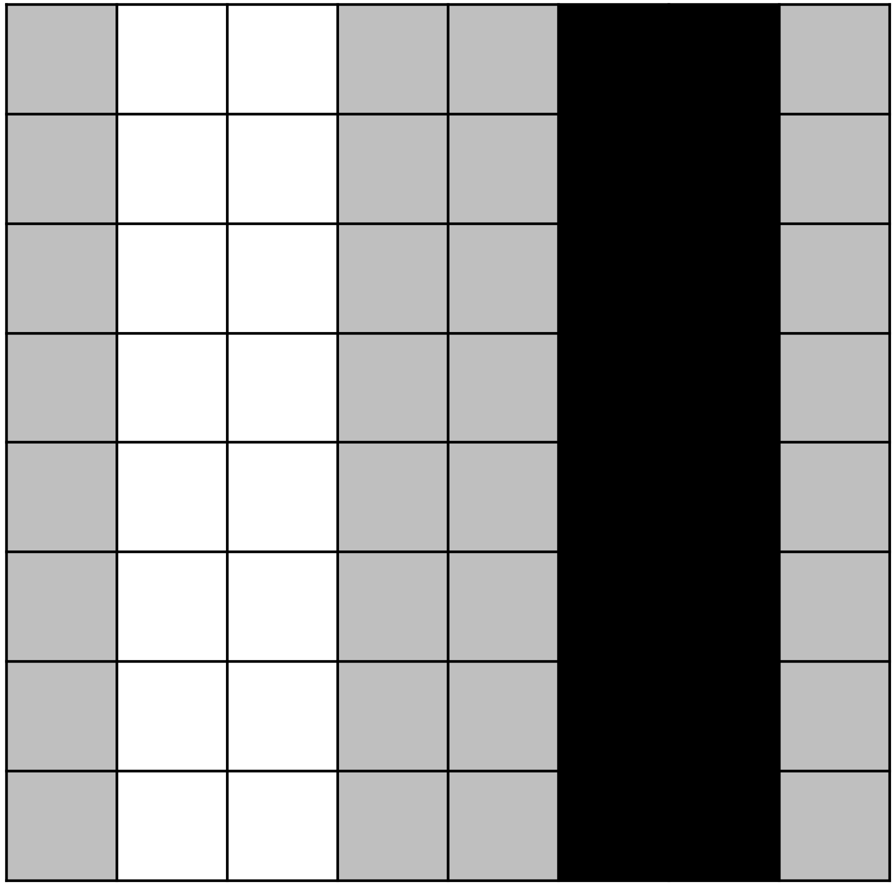
0 3 3 0 0 -3 -3 0

-1	0	1
-1	0	1
-1	0	1

Convolution



0 3 3 0 0 -3 -3 0



-1	0	1
-1	0	1
-1	0	1

Convolution: Padding

0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0
0	0	1	1	1	1	0	0

0	2	2	0	0	-2	-2	0
0	3	3	0	0	-3	-3	0
0	3	3	0	0	-3	-3	0
0	3	3	0	0	-3	-3	0
0	3	3	0	0	-3	-3	0
0	3	3	0	0	-3	-3	0
0	3	3	0	0	-3	-3	0
0	2	2	0	0	-2	-2	0

Quiz: Which kernel goes with which output image?

1

-1	0	1
-2	0	2
-1	0	1

2

-1	-2	-1
0	0	0
1	2	1

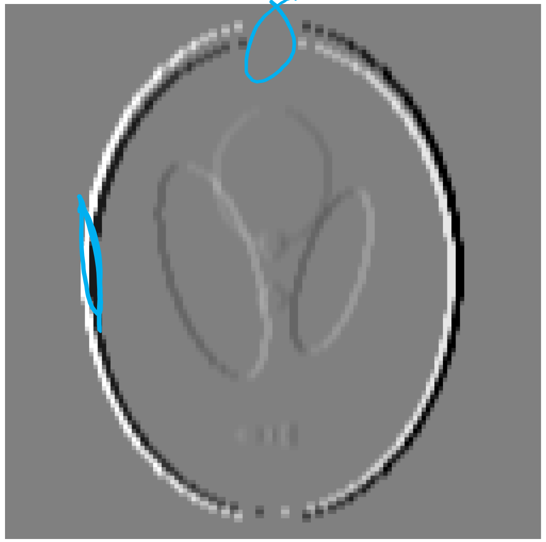
3

0	0	-1	0
0	-2	0	1
-1	0	2	0
0	1	0	0

A

B

C



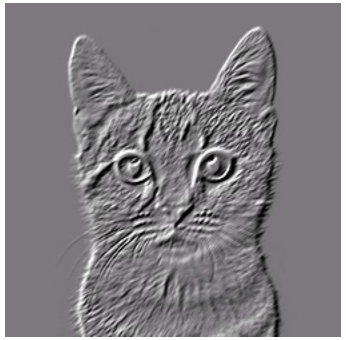
Convolutional Neural Networks

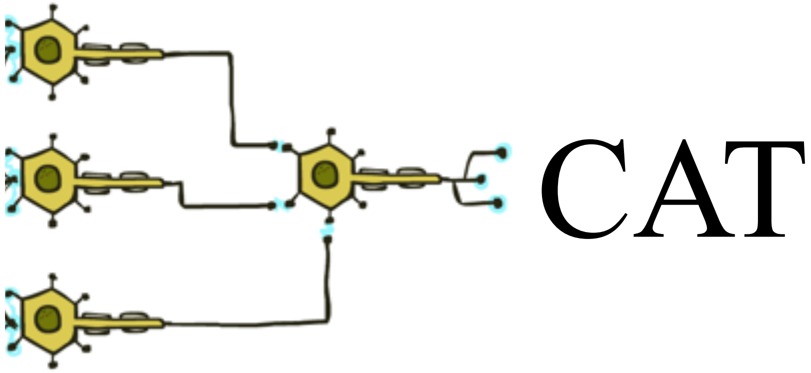
Convolution

-1	0	1
-2	0	2
-1	0	1

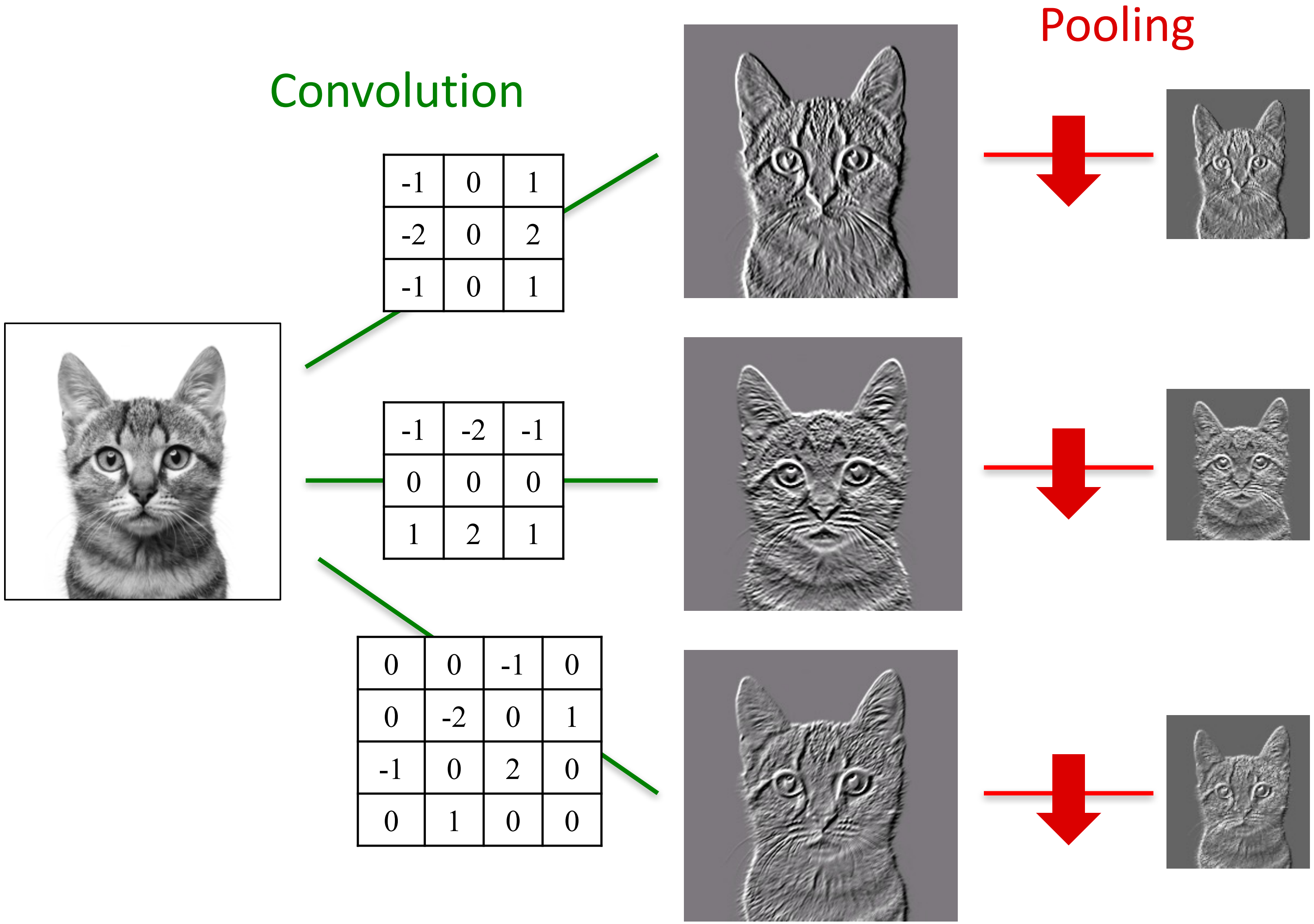
-1	-2	-1
0	0	0
1	2	1

0	0	-1	0
0	-2	0	1
-1	0	2	0
0	1	0	0





Convolutional Neural Networks



Convolution: Stride=2

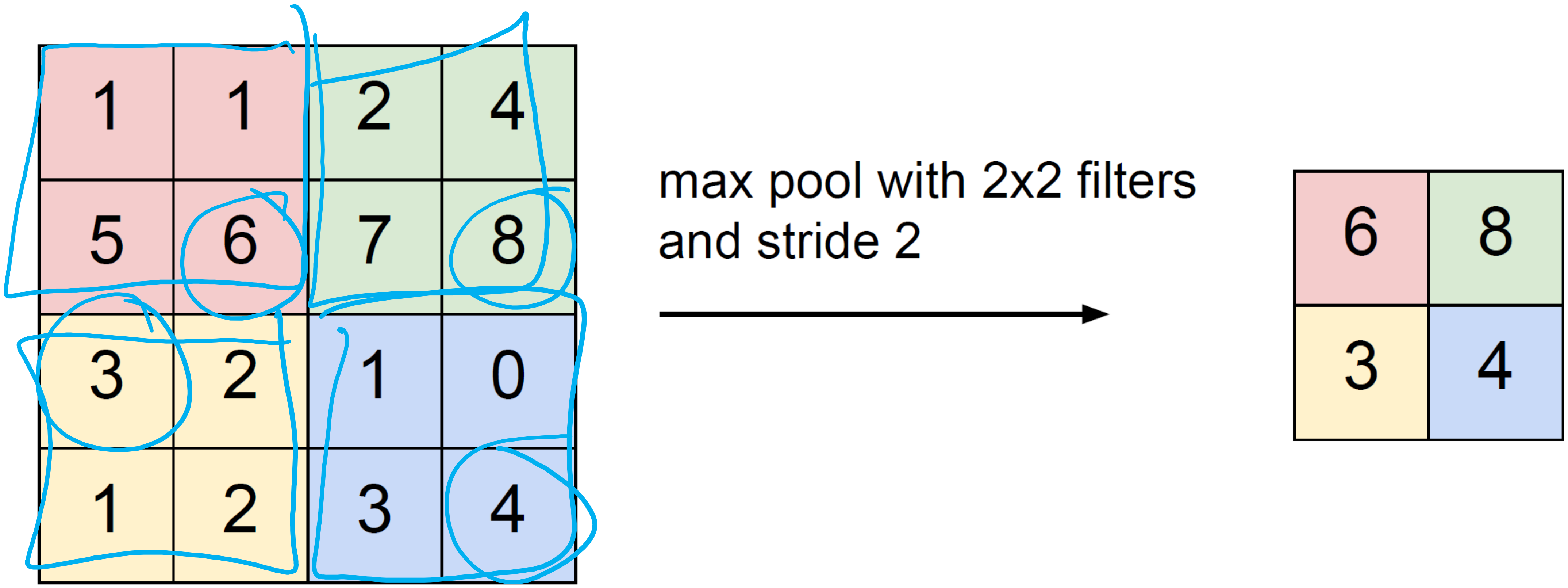
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	0

.25	.25
.25	.25

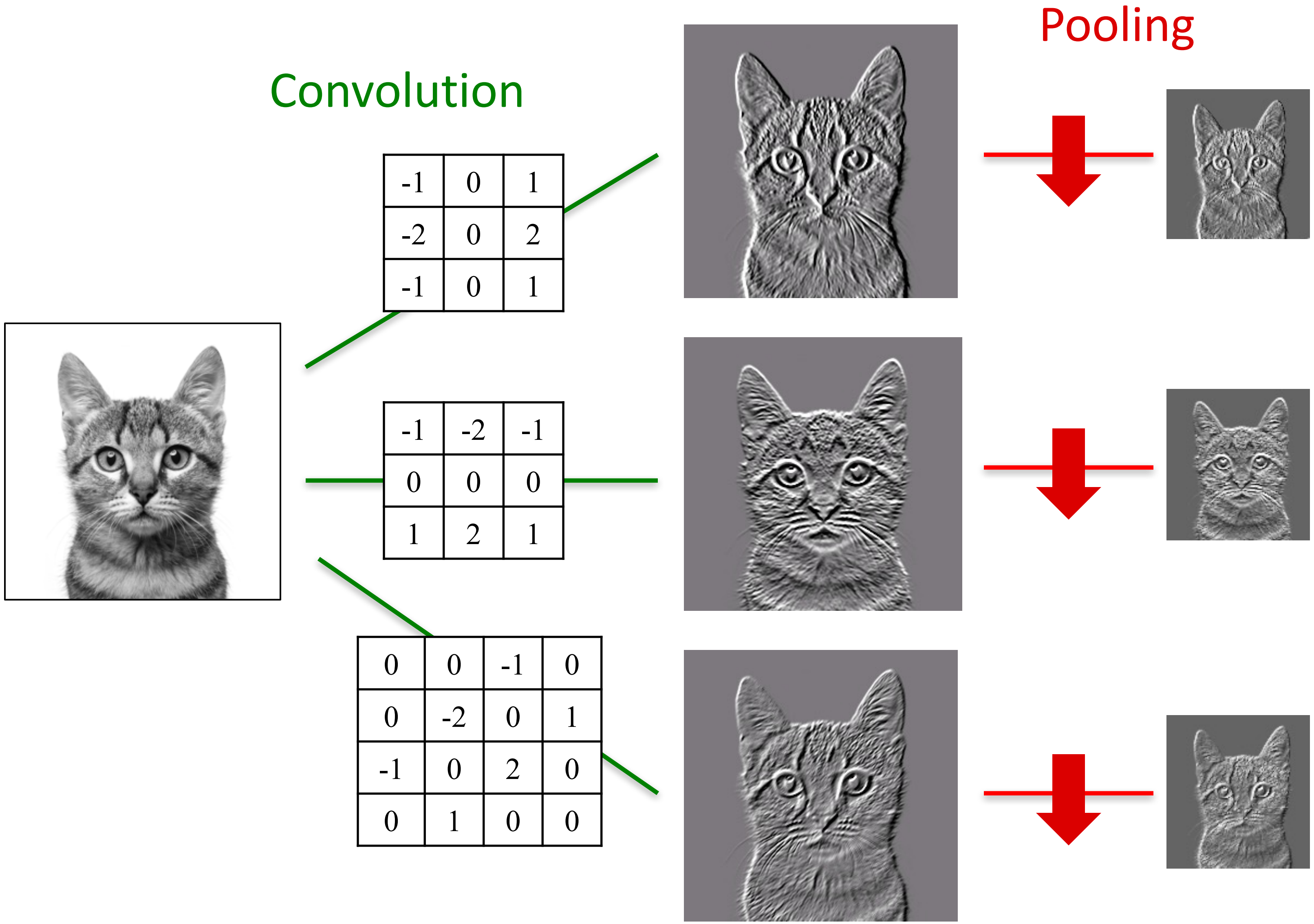
0 .5 1 .5 0

0 .5 1 .5 0

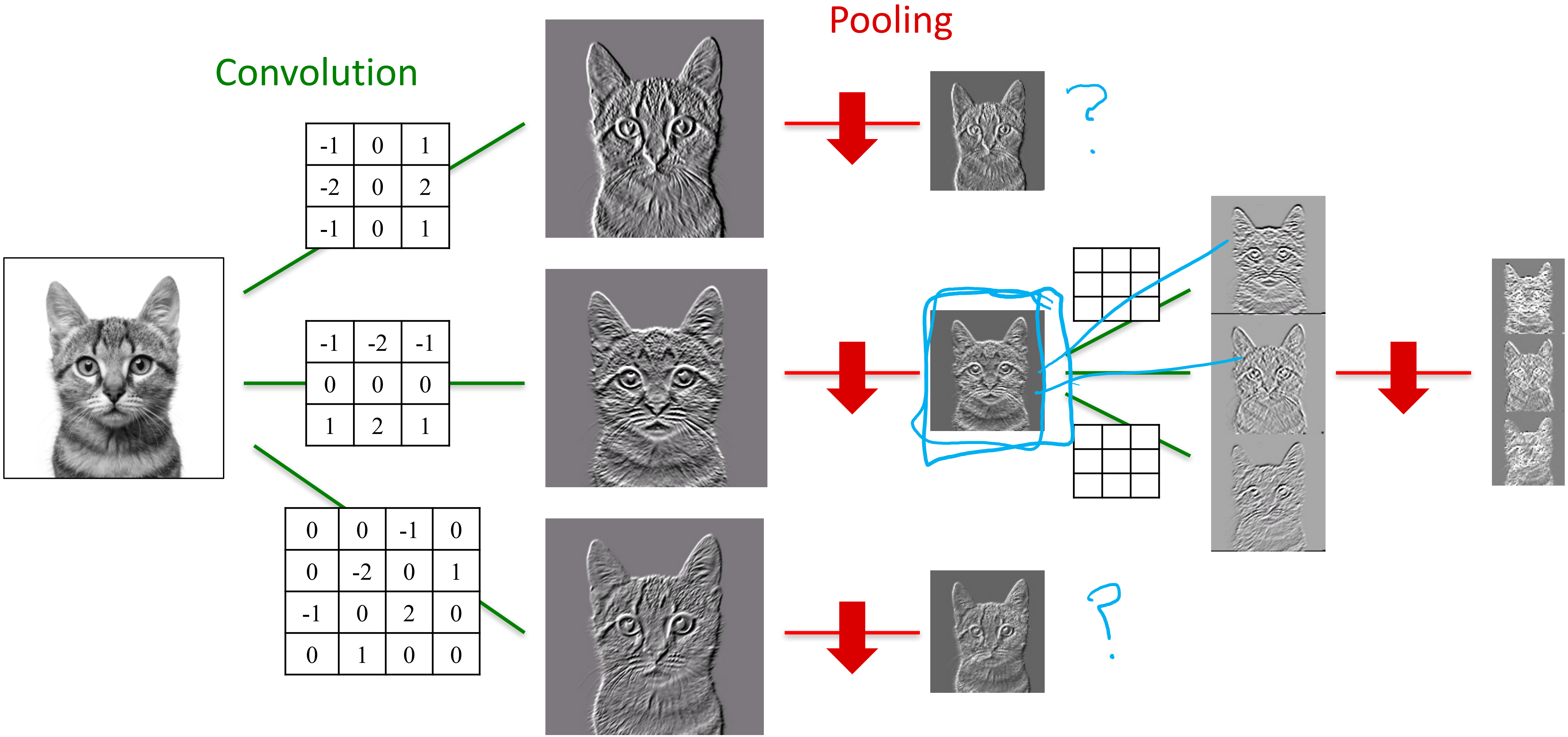
Stride: Max Pooling



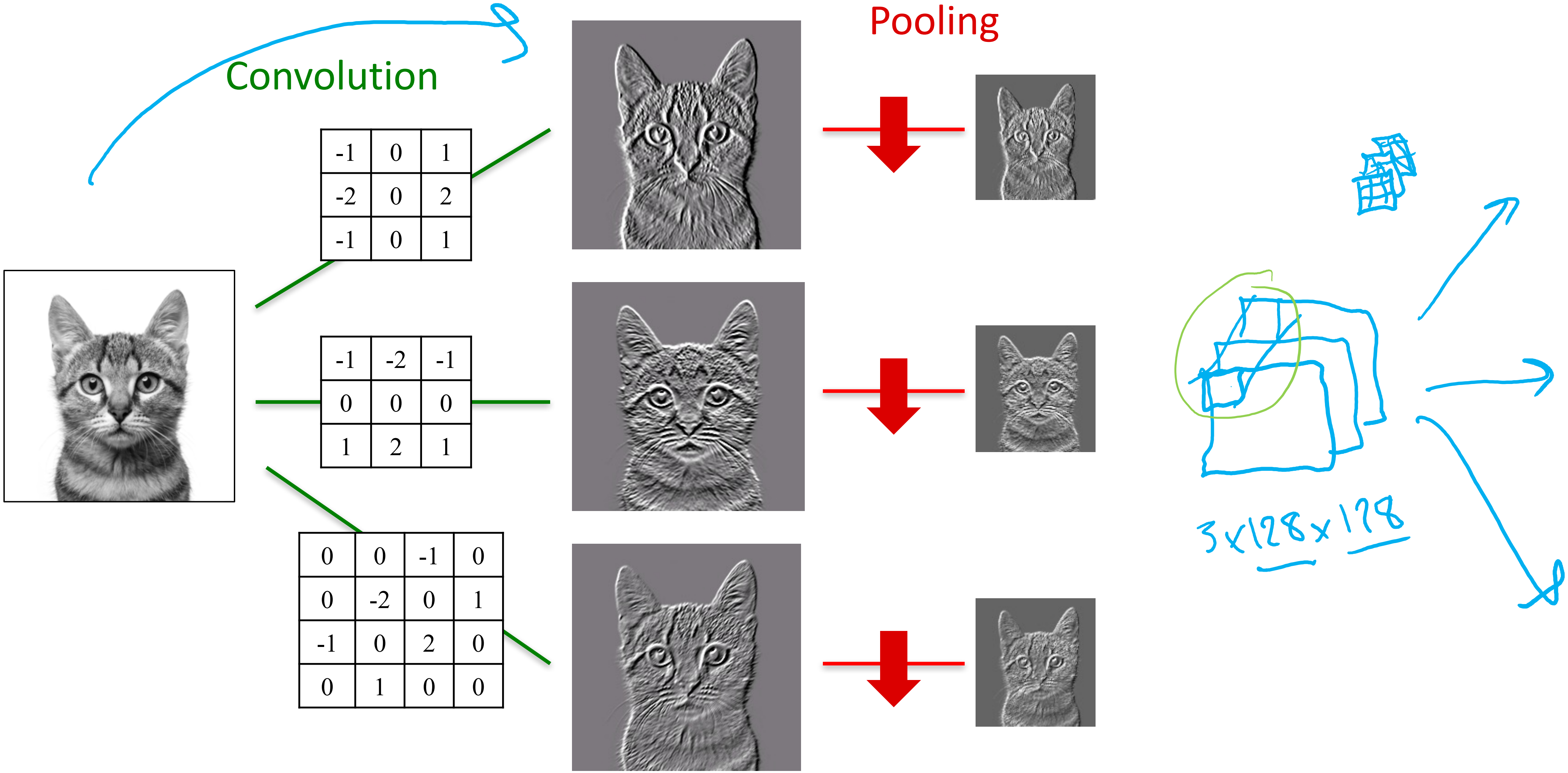
Convolutional Neural Networks



Convolutional Neural Networks

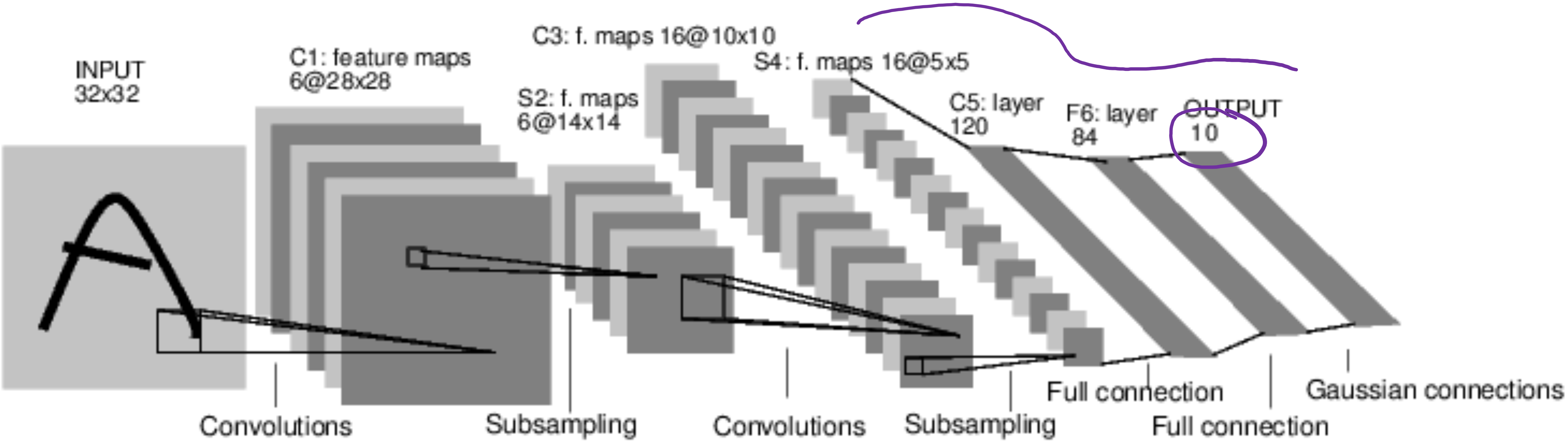


Convolutional Neural Networks



Convolutional Neural Networks

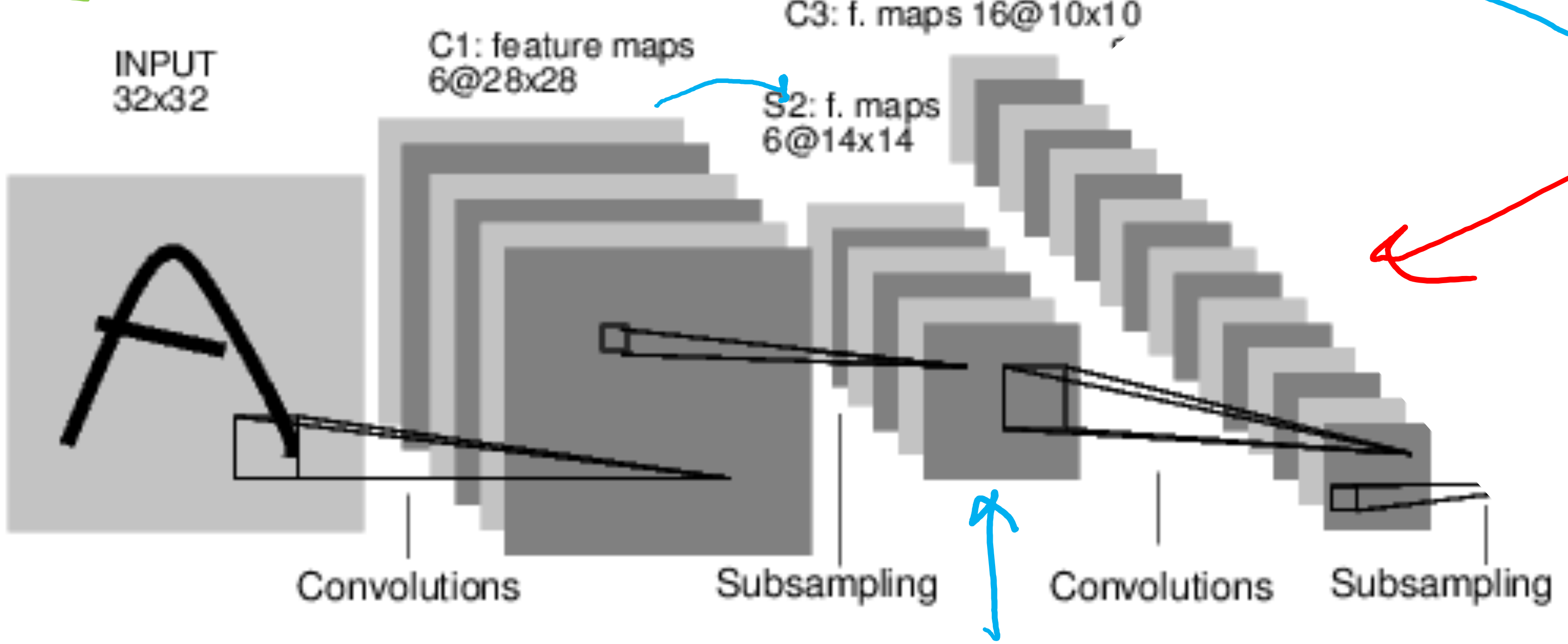
- Lenet5 – Lecun, et al, 1998
  - Convnets for digit recognition

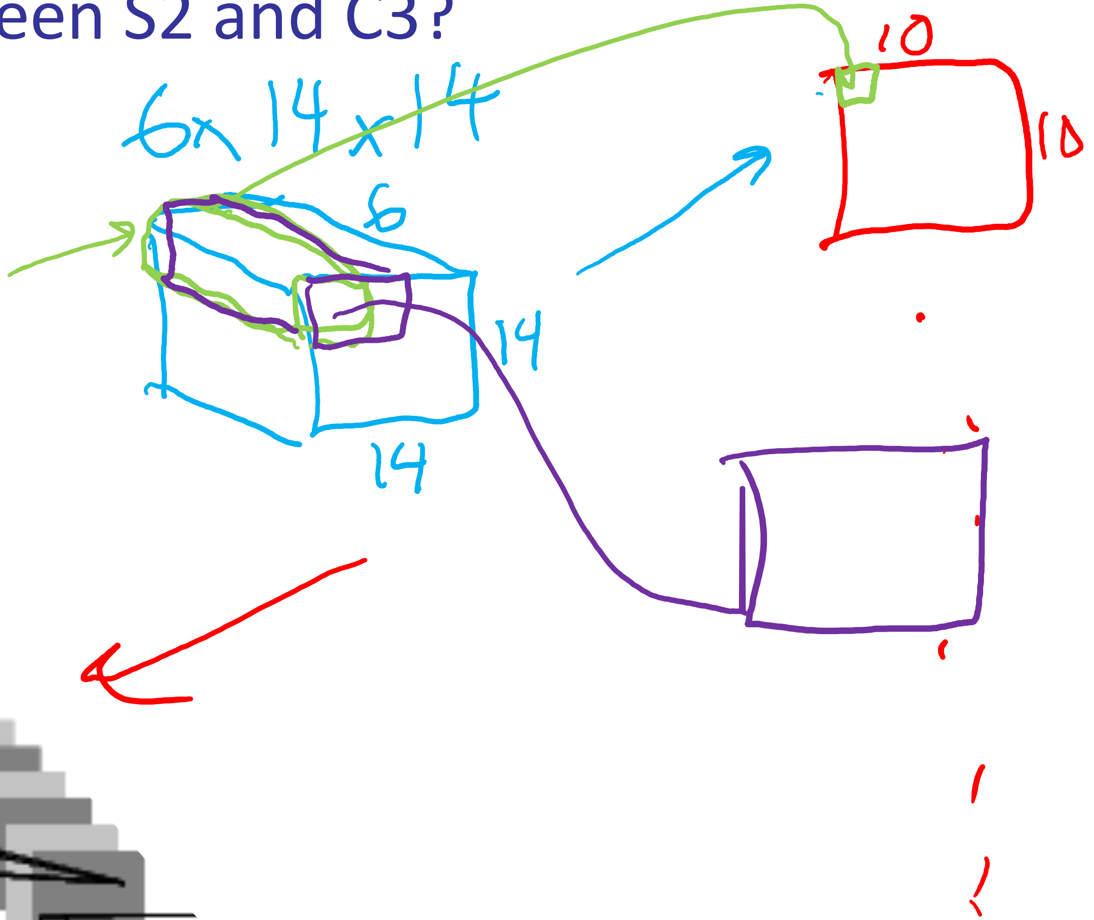


LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

Quiz: How many weights?

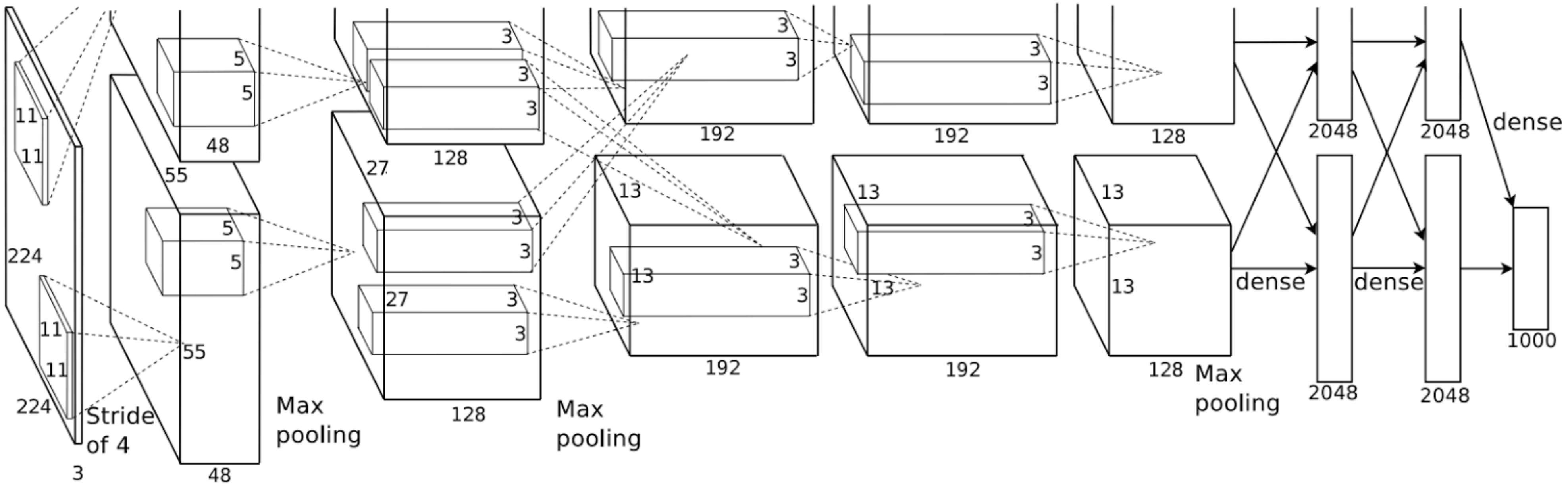
- How big many convolutional weights between S2 and C3?
  - S2: 6 channels @14x14
  - Conv: 5x5, pad=1, stride=1
  - C3: 16 channels @ 10x10





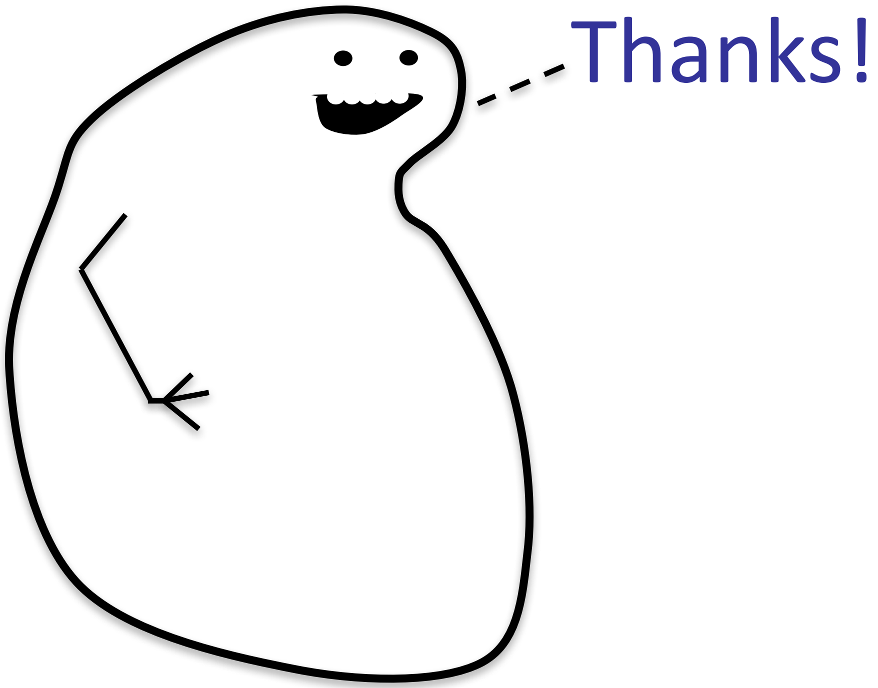
Convolutional Neural Networks

- Alexnet – Lecun, et al, 2012
  - Convnets for image classification
  - More data & more compute power



Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." NIPS, 2012.

That's All Folks



Pat Virtue
virtue@eecs.berkeley.edu