INTRODUCTION TO MACHINE LEARNING COMPSCI 4ML3

LECTURE 24

HASSAN ASHTIANI

NO FREE LUNCH

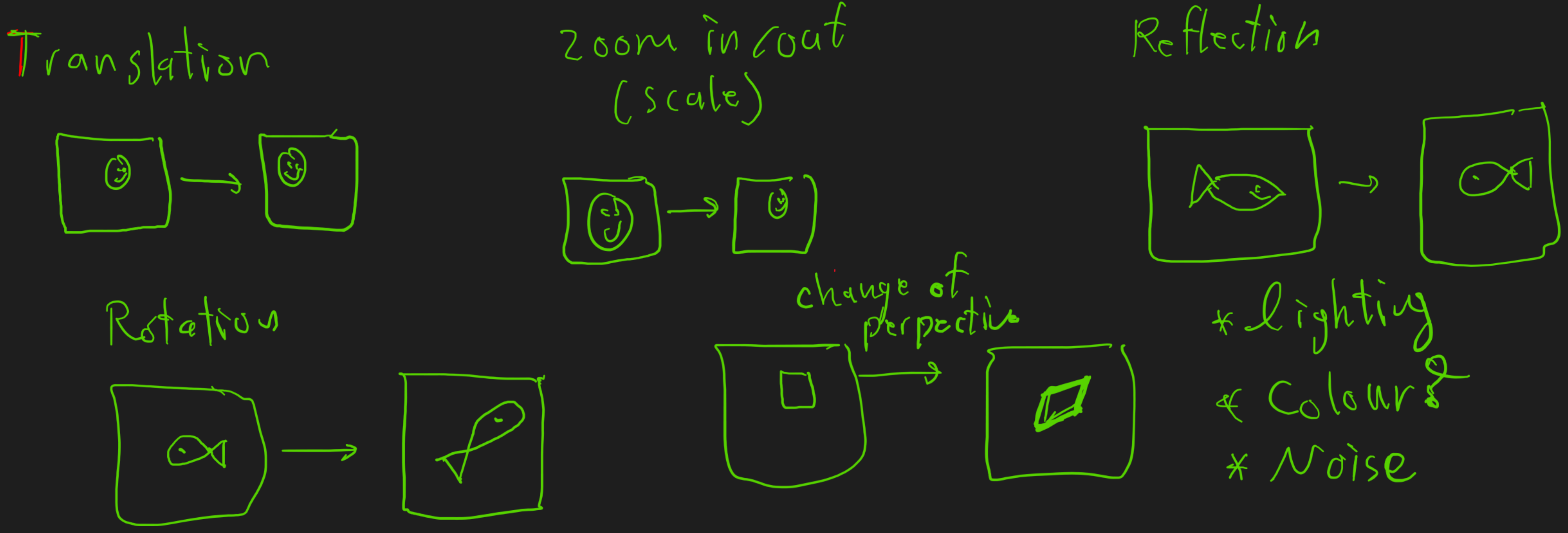
- Larger networks can fit any data set but
  - the more flexible the model the more training data is required
  - Computational complexity

How can we address these drawbacks?

  - Incorporate domain knowledge

THE CASE OF IMAGE CLASSIFICATION

- Image data has a lot of "structure"
  - Invariance

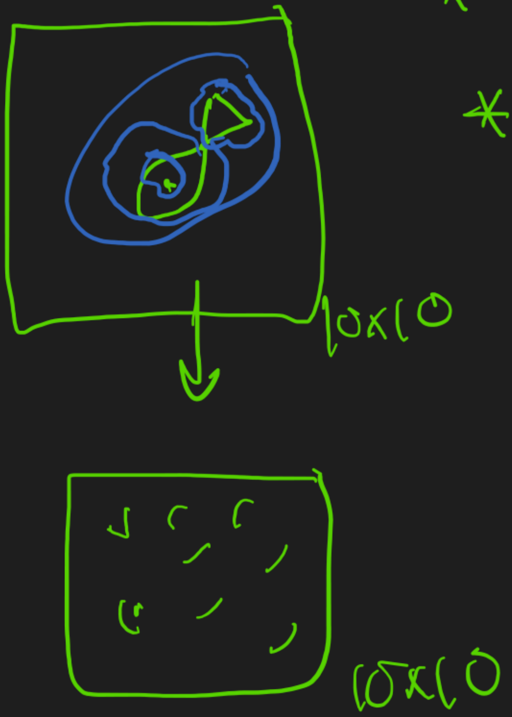


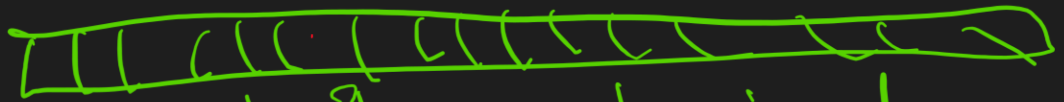
THE CASE OF IMAGE CLASSIFICATION

- Image data has a lot of "structure"
  - Locality

* spatial relationships

* nearby pixels are correlated





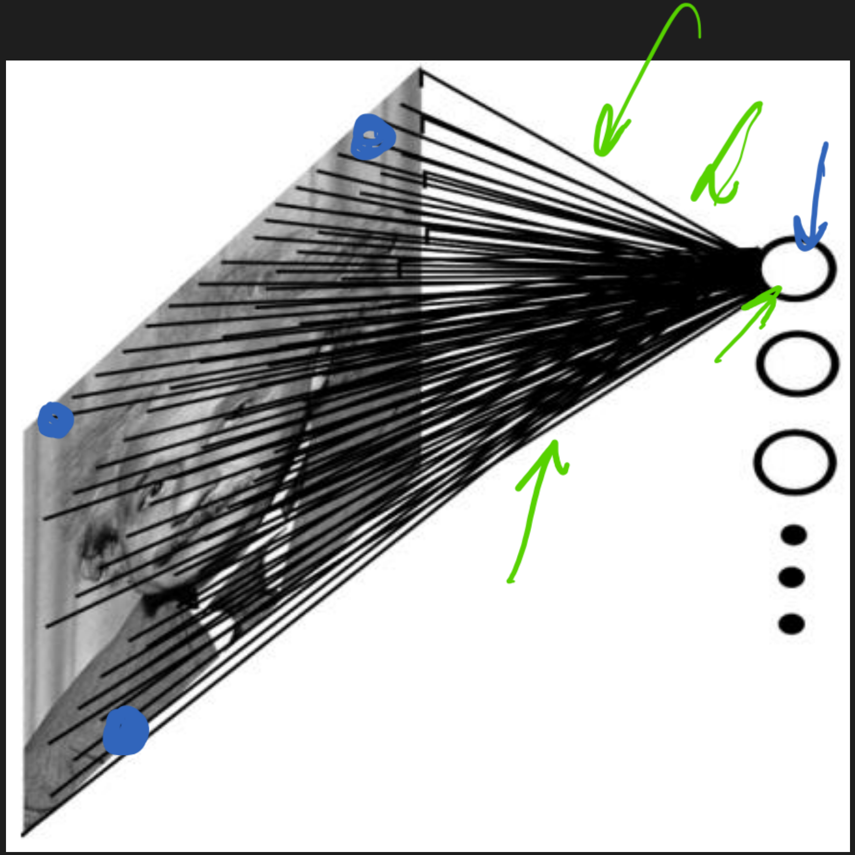
don't vectorize!

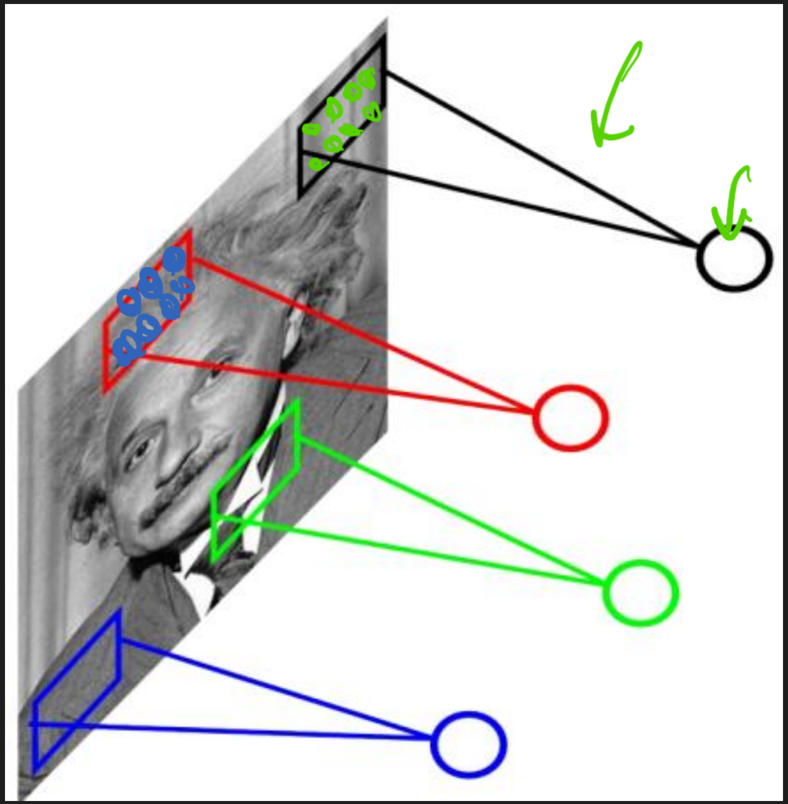
STRUCTURAL REGULARIZATION

- Can we use the properties of images to reduce the number of parameters, without compromising the discriminative power of them?
- Regularization by literally reducing the number of parameters

EXPLOITING LOCALITY

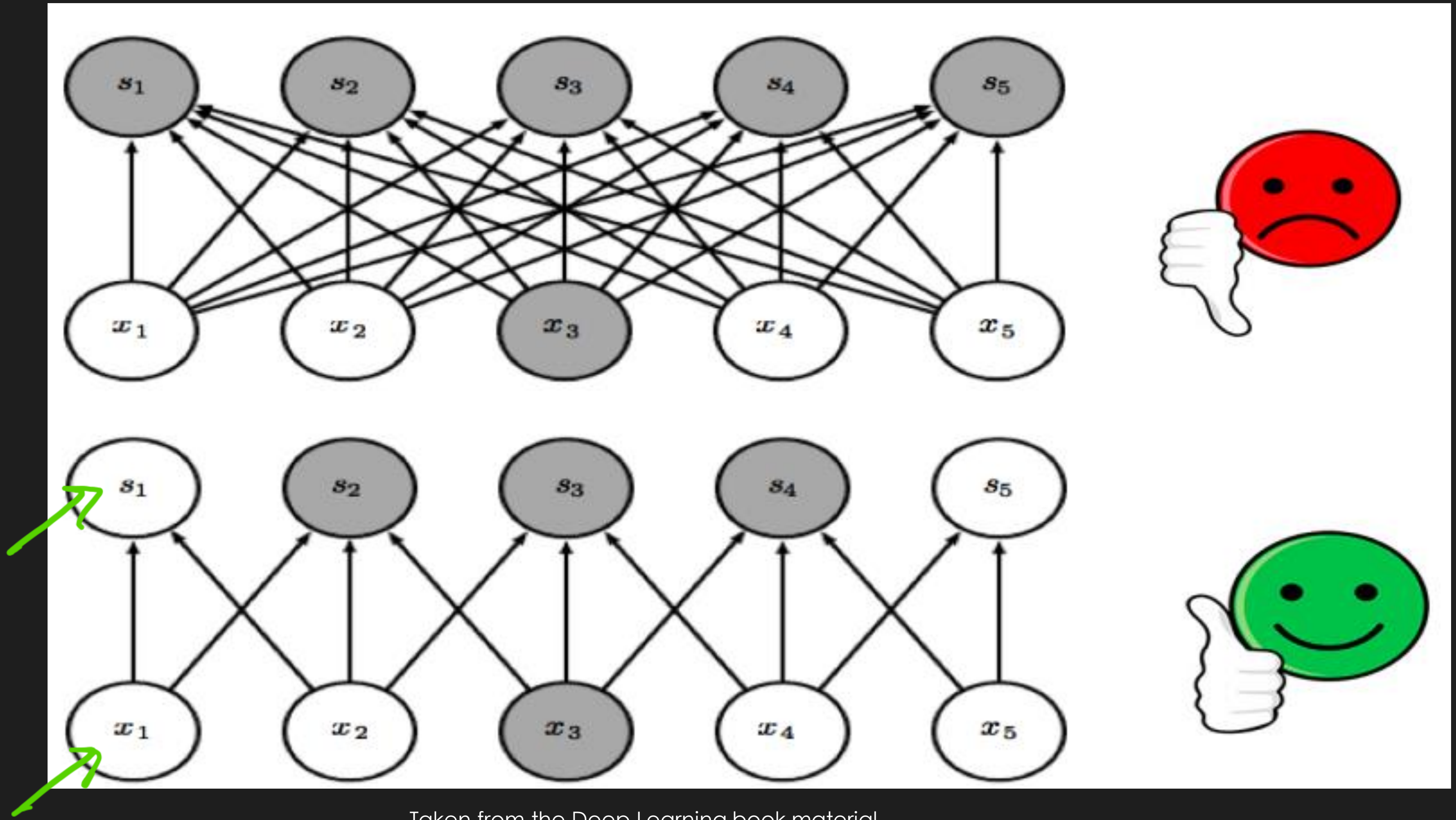
- Sparse connectivity rather than full connectivity





Pictures taken from Ranzato slides

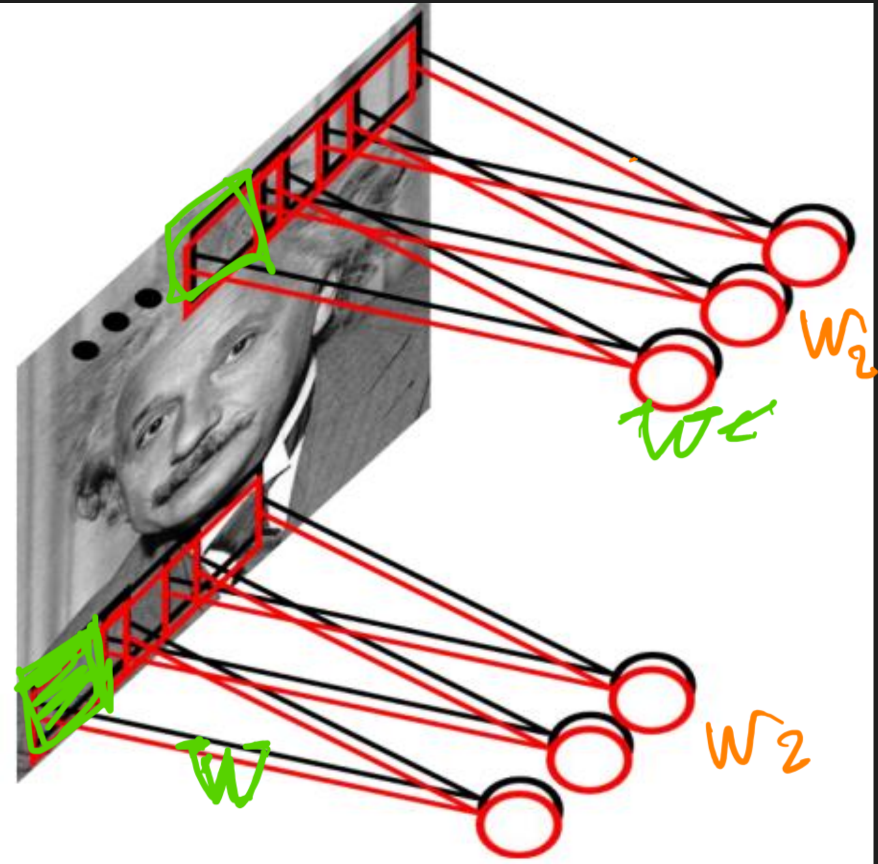
EXPLOITING LOCALITY



Taken from the Deep Learning book material

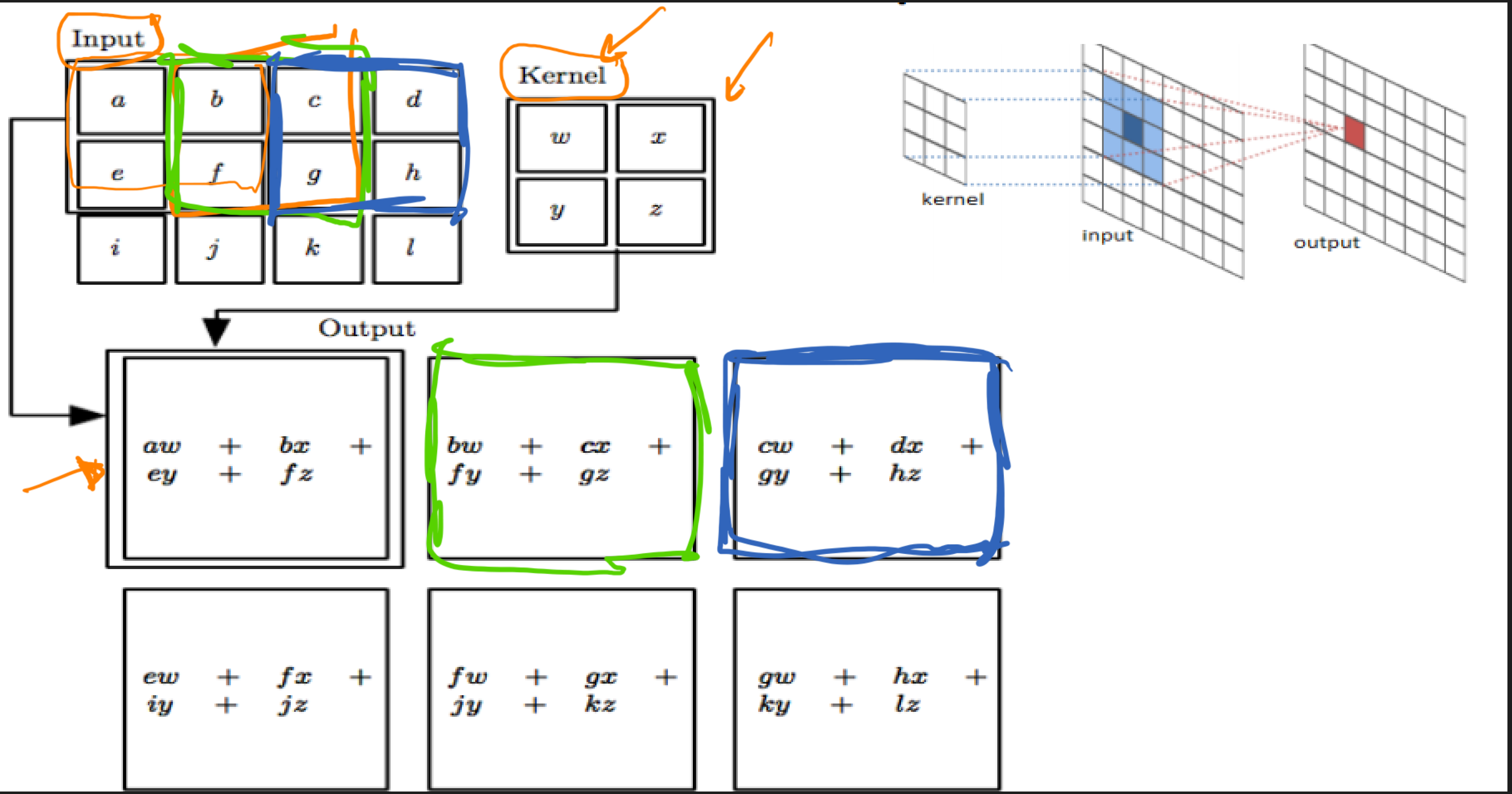
EXPLOITING INVARIANCE

- If extracting a nose is useful in one part of the image, it will be useful in other parts of the image as well
- Parameter sharing!



THE CONVOLUTION OPERATOR

To be more accurate, the kernel should be flipped.

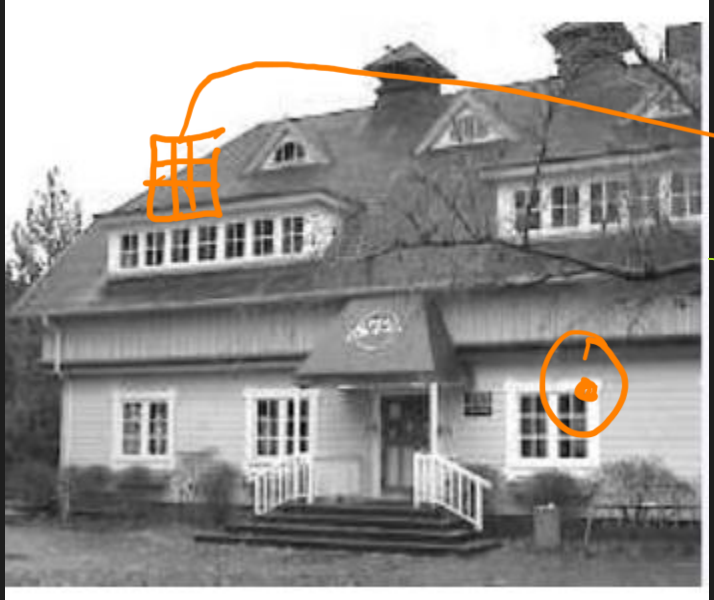


Taken from the Deep Learning book material

CONVOLUTION OPERATOR

- 1-D CONVOLUTION
  - $y = (x * w)$
  - $y(i) = \sum_t x(t)w(i - t)$
- 2-D CONVOLUTION
  - $y = (x * w)$
  - $y(i, j) = \sum_{t_1} \sum_{t_2} x(t_1, t_2)w(i - t_1, j - t_2)$
- USEFUL NOT ONLY FOR IMAGES, BUT FOR OTHER SIGNALS

BLUR



Kernel

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

Example taken from http://aishack.in/tutorials/image-convolution-examples/

GAUSSIAN BLUR

0	0	0	5	0	0	0
0	5	18	32	18	5	0
0	18	64	100	64	18	0
5	32	100	100	100	32	5
0	18	64	100	64	18	0
0	5	18	32	18	5	0
0	0	0	5	0	0	0

HORIZONTAL LINE

-1	-1	-1
2	2	2
-1	-1	-1

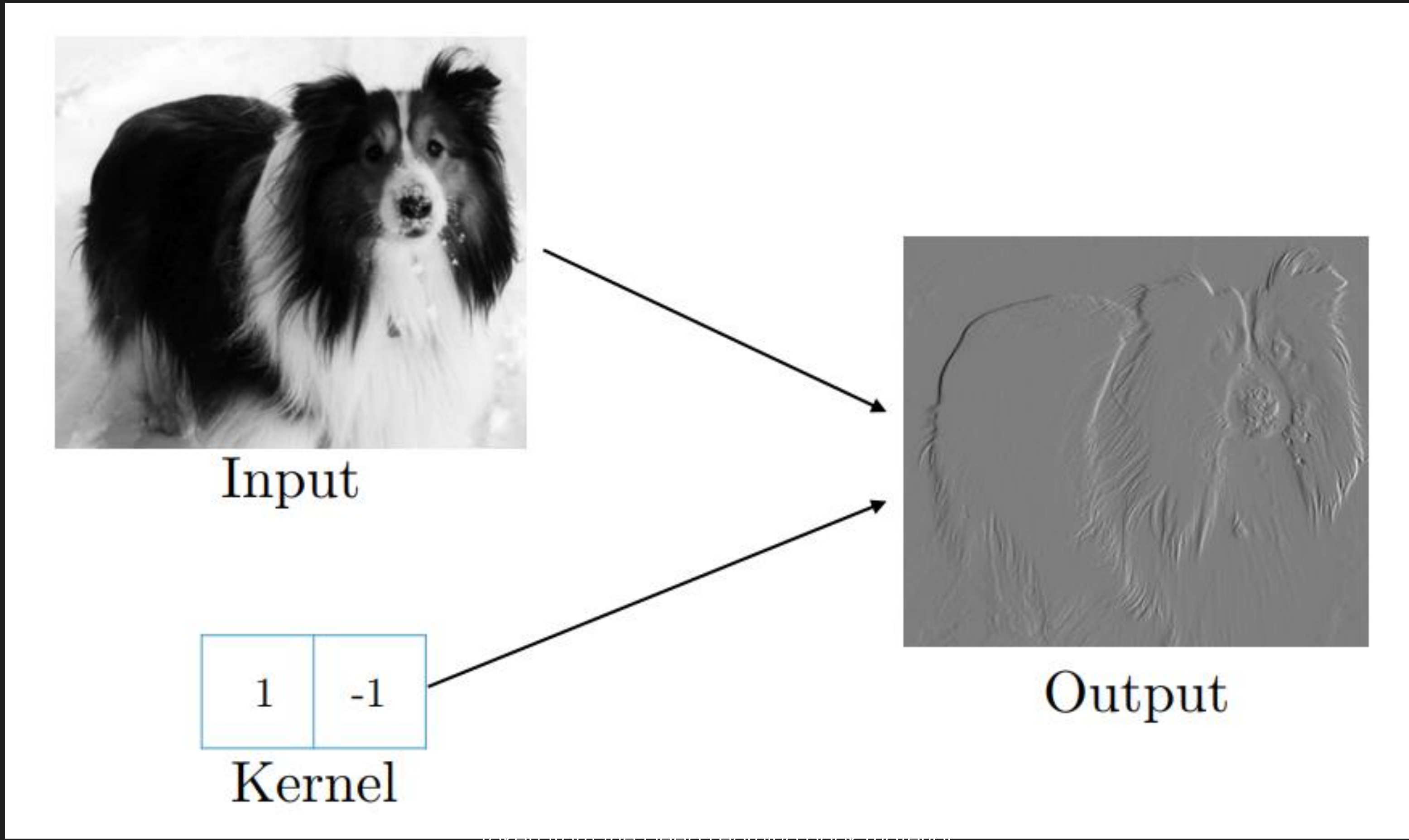
A KIND OF EDGE DETECTOR

-1	-1	-1
-1	8	-1
-1	-1	-1

A KIND OF EDGE DETECTOR

THE USE OF CONVOLUTION FOR IMAGE PROCESSING IS QUITE OLD

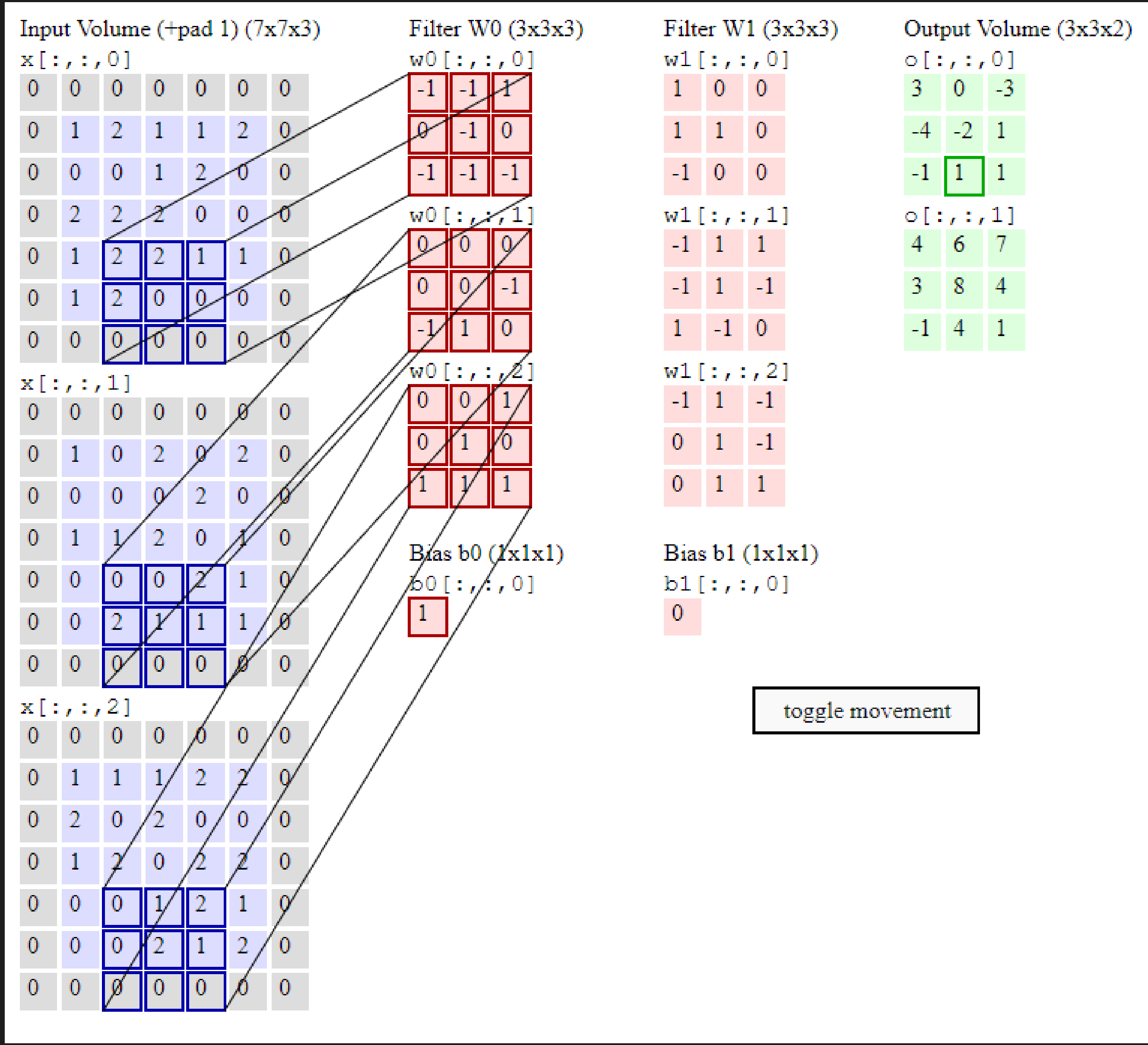
..BUT USING AN END-TO-END LEARNING APPROACH WHERE THE FILTERS/KERNELS ARE ALSO LEARNED IS THE POWER OF CONVOLUTIONAL NEURAL NETWORKS



Taken from the Deep Learning book material

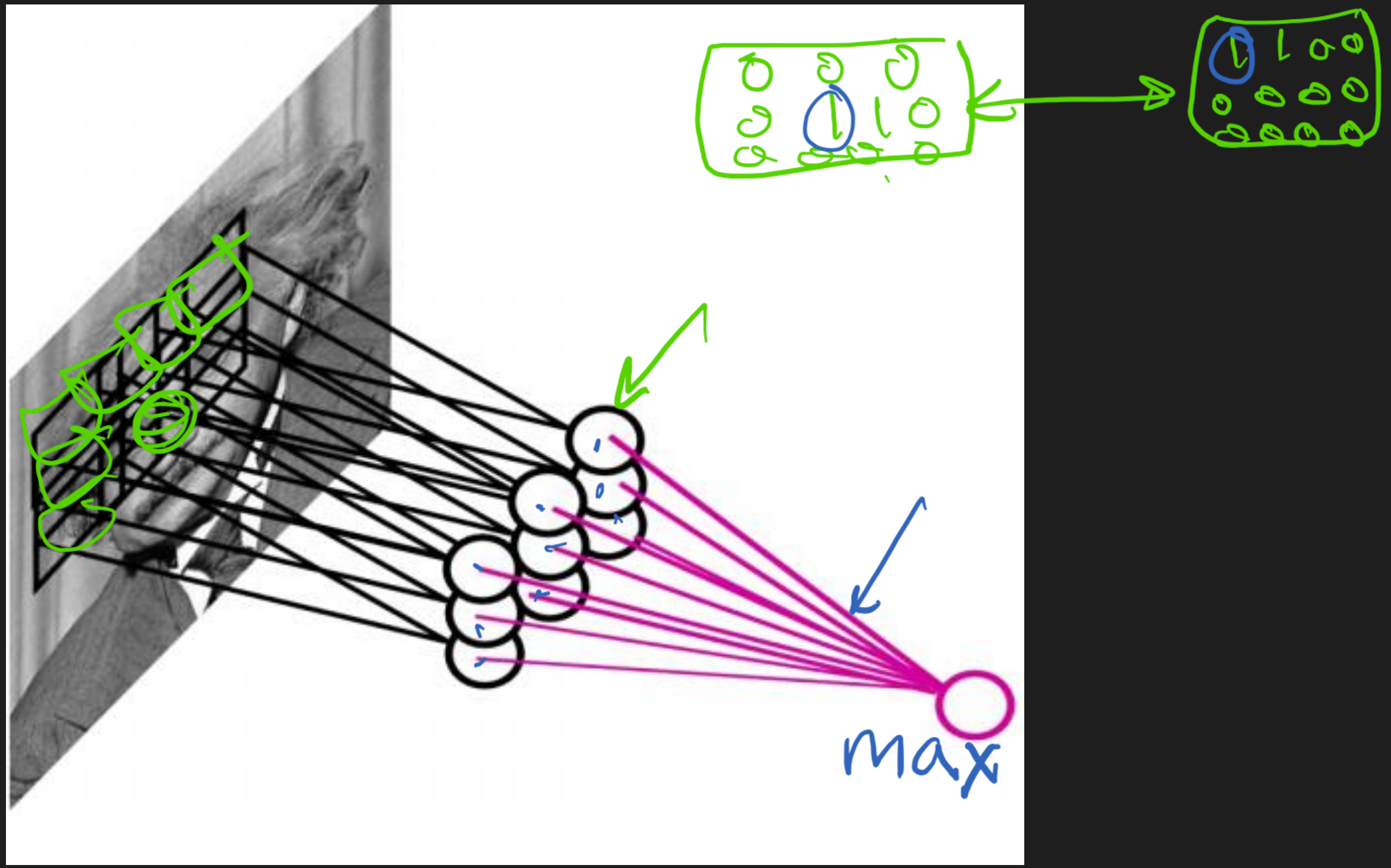
HTTP://CS231N.GITHUB.IO/CONVOLUTIONAL-NETWORKS/

- PADDING
- STRIDE
- CHANNEL
- KERNEL VS FILTER



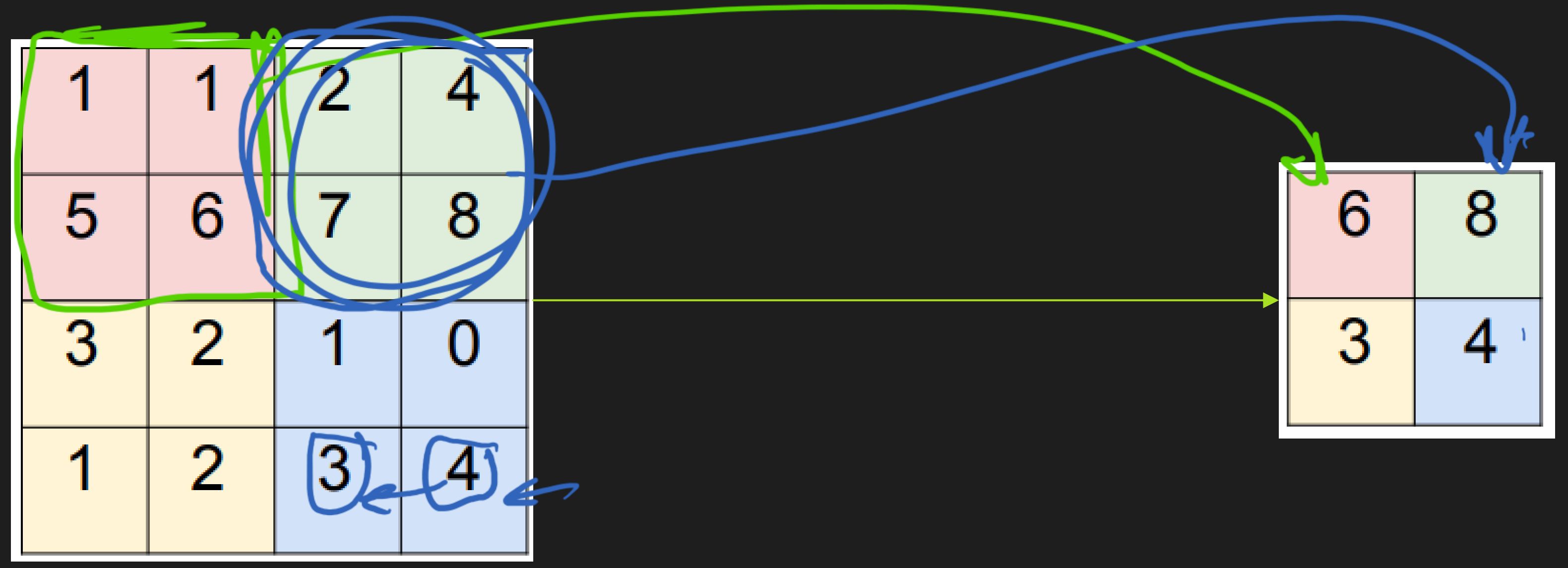
EXPLOITING LOCAL TRANSLATION-INVARIANCE

- IT DOES NOT MATTER EXACTLY WHICH OF THE SMALL PATCHES OF THE IMAGE INCLUDE A NOSE!
- MAX POOLING LAYER REDUCES THE NUMBER OF PARAMETERS



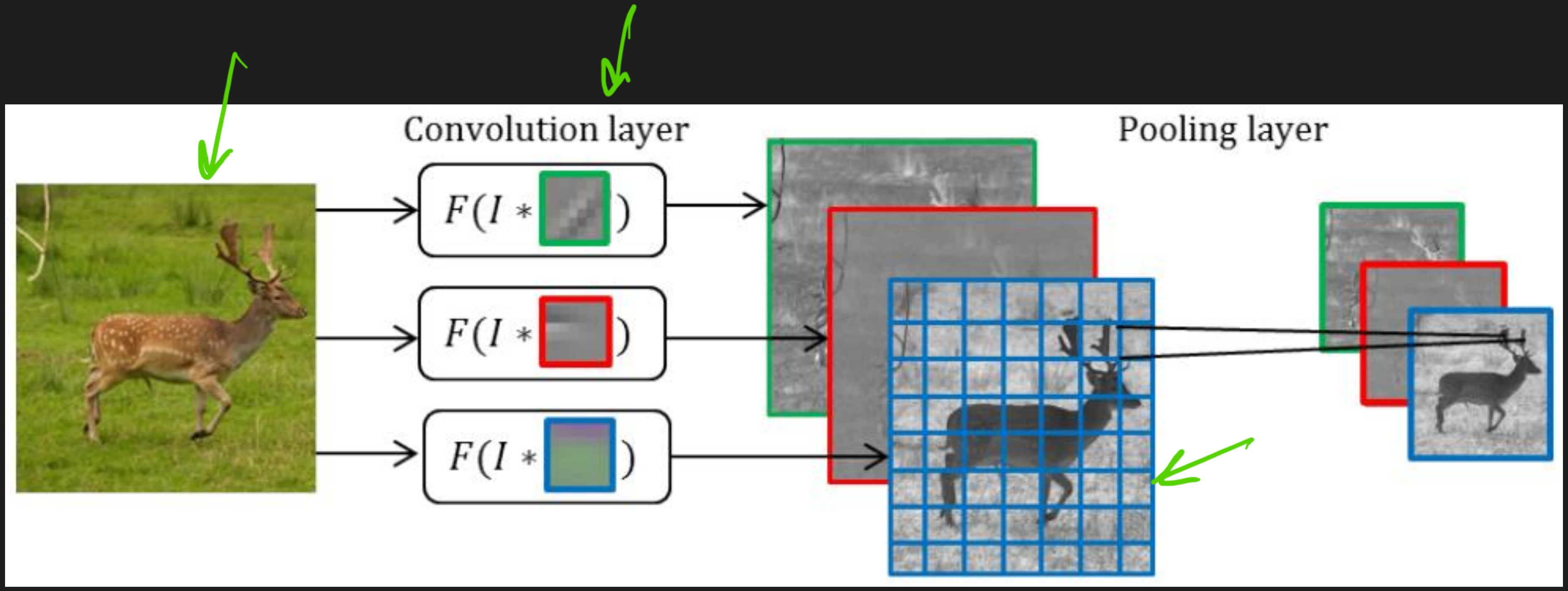
MAX POOLING

- 2x2 max pooling
  - Stride=2, no padding

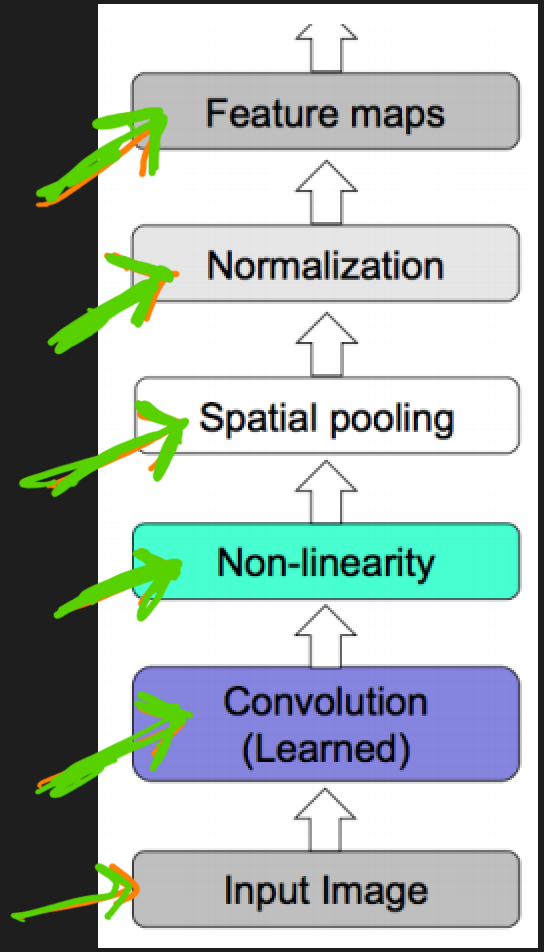


  - Stride=1?

A FULL CONVOLUTIONAL LAYER

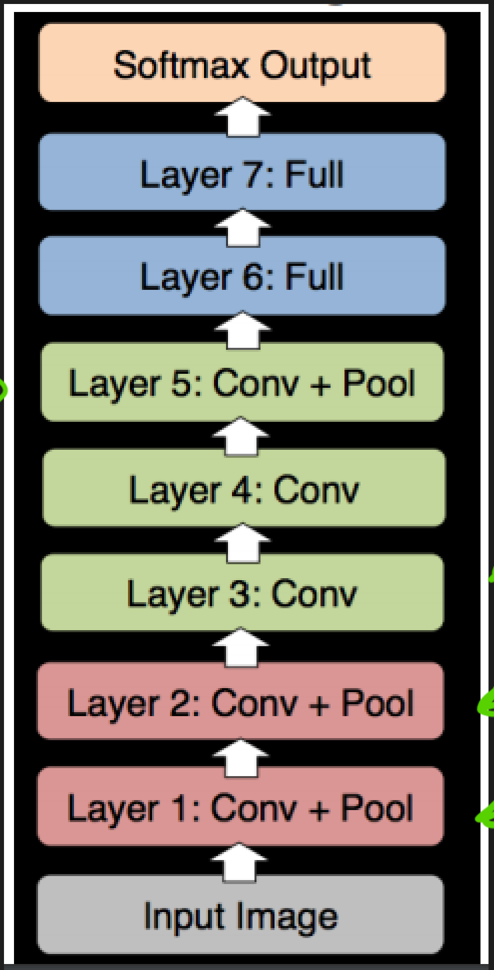


Taken from Barshan et al., "SCALABLE MULTI-NEIGHBORHOOD LEARNING FOR CONVOLUTIONAL NETWORKS"



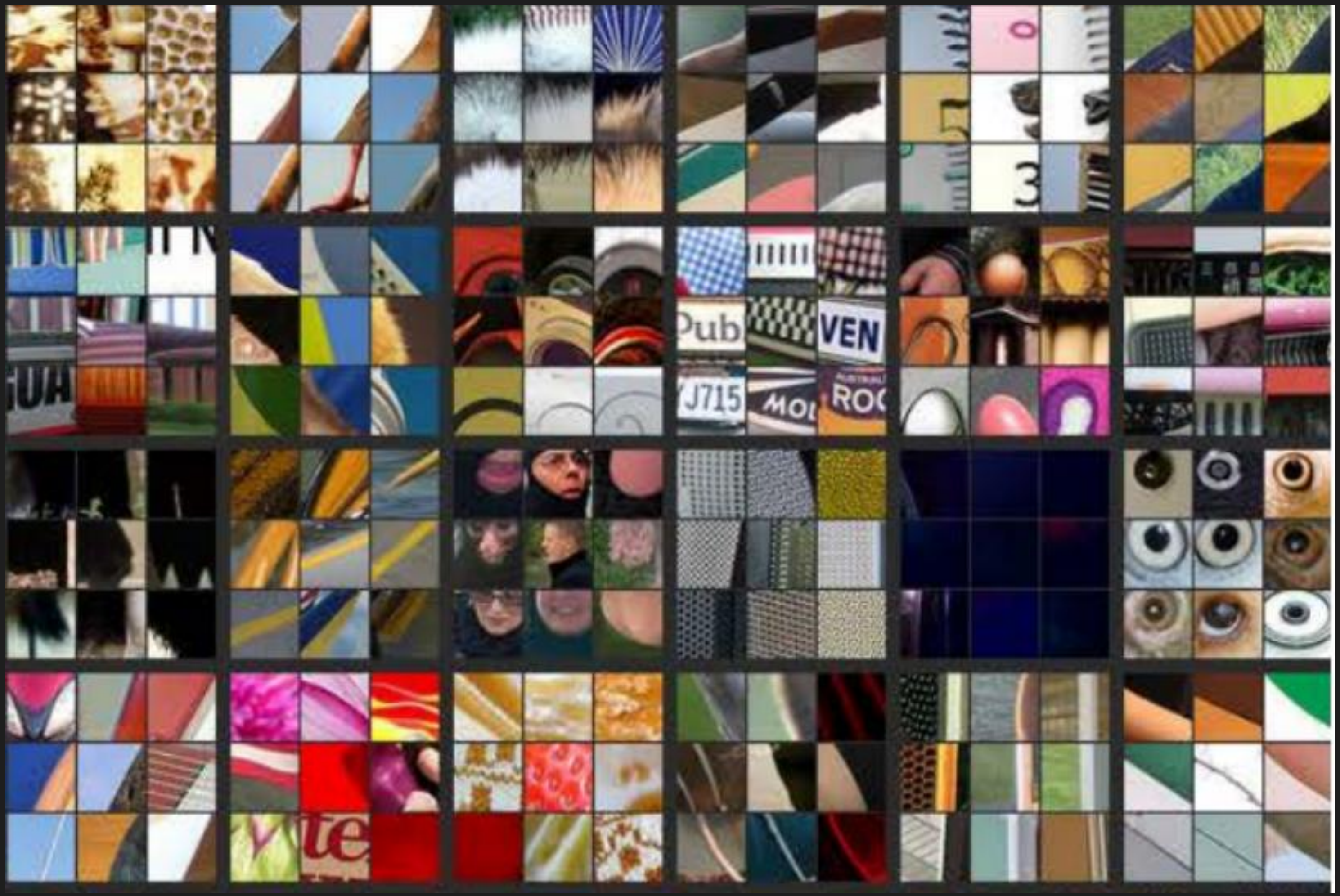
BREAKTHROUGH IN IMAGENET

- ImageNet classification challenge
  - Millions of images
  - Thousands of classes
- In 2012, AlexNet used won the competition by a high margin
  - ~15% error compared to ~25% of the next team
  - They used a convolutional architecture
  - They used GPUs for speedup
- CNNs became very popular



THINGS THE FIRST LAYER DETECTS

THINGS THE 2ND LAYER DETECTS



3RD LAYER

5TH LAYER