ANNOUNCEMENTS -

NOT OPTIONAL / EXTRA-CREDIT!

1. P8 released, due 05/04, late due 05/06 – last day of classes
2. [Extra-credit] Kaggle, HW8 to be released

OTHER ANNOUNCEMENTS –

1. Pre-enrollment starts Monday (04/28)

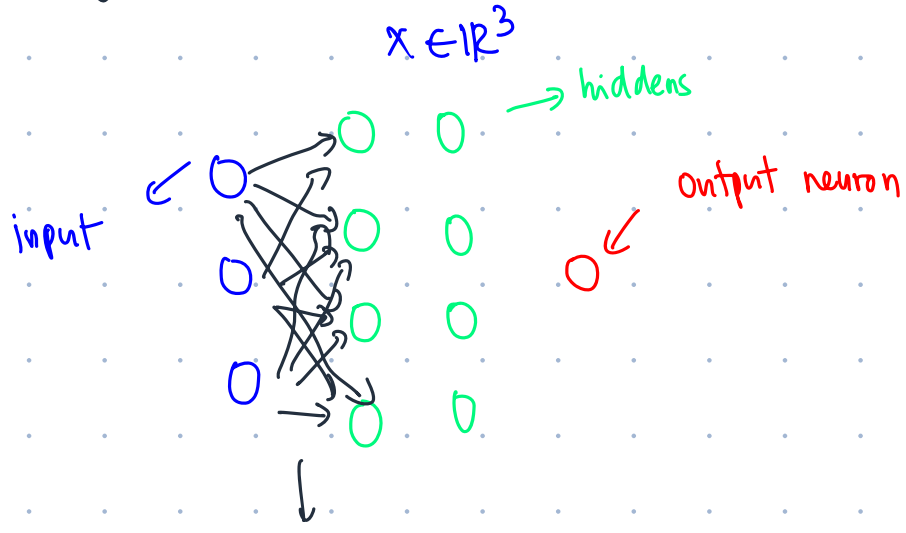
   Possible next steps to 3780 –

CS 3700	Foundations of AI Reasoning and Decision-Making
CS 4750	Foundation of Robotics
CS 4756	Robot Learning
CS 4782	Deep Learning
CS 4670	Introduction to Computer Vision
CS 4789	Introduction to Reinforcement Learning
CS 4783	Mathematical Foundations of Machine Learning
CS 4740	Natural Language Processing
CS 4787	Principles of Large-Scale Machine Learning Systems

2. Want to TA for 3780/5780 next semester, apply soon!

Last lecture + last week –

– Fully-connected NN (FCN)



how many lines did I draw? – 12 lines, OR 12 weights

$$x^{(l)} = g\left(W^{(l)} x^{(l-1)} + b^{(l)}\right)$$

↑ non-linearity

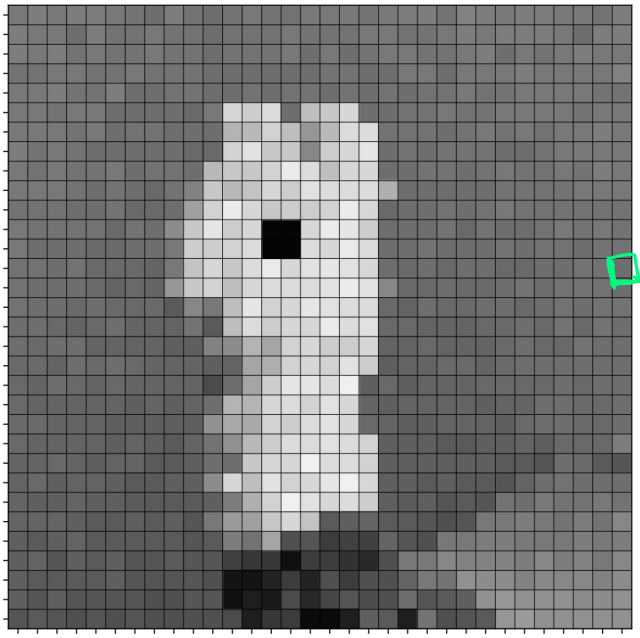
$x^{(l-1)} \in \mathbb{R}^3$

$W^{(l)} \in \mathbb{R}^{4\times 3}$ (Rahim)

$b^{(l)} \in \mathbb{R}^4$ (Aditya)

Same model, different modality

let's start simple, just grayscale!



We want to use FCN to classify Berry!

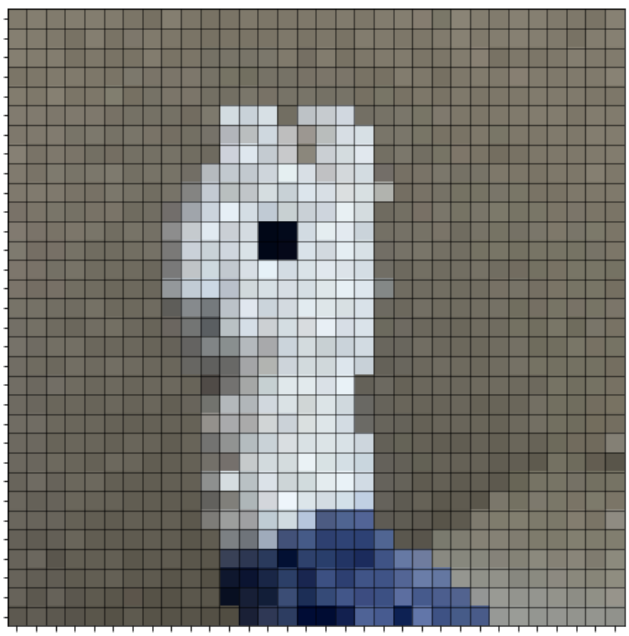
Dom says

→ b/w 0,1 ⇒ 0.0 = black, 1.0 = white } 0.6 | 0.7

Vanija says, 32x32 = 1024 pixels

→ [0.6 ⋮ 0.9] → 1024-dim

O O ⋮ O hidden → 1024 weights per neuron per layer!

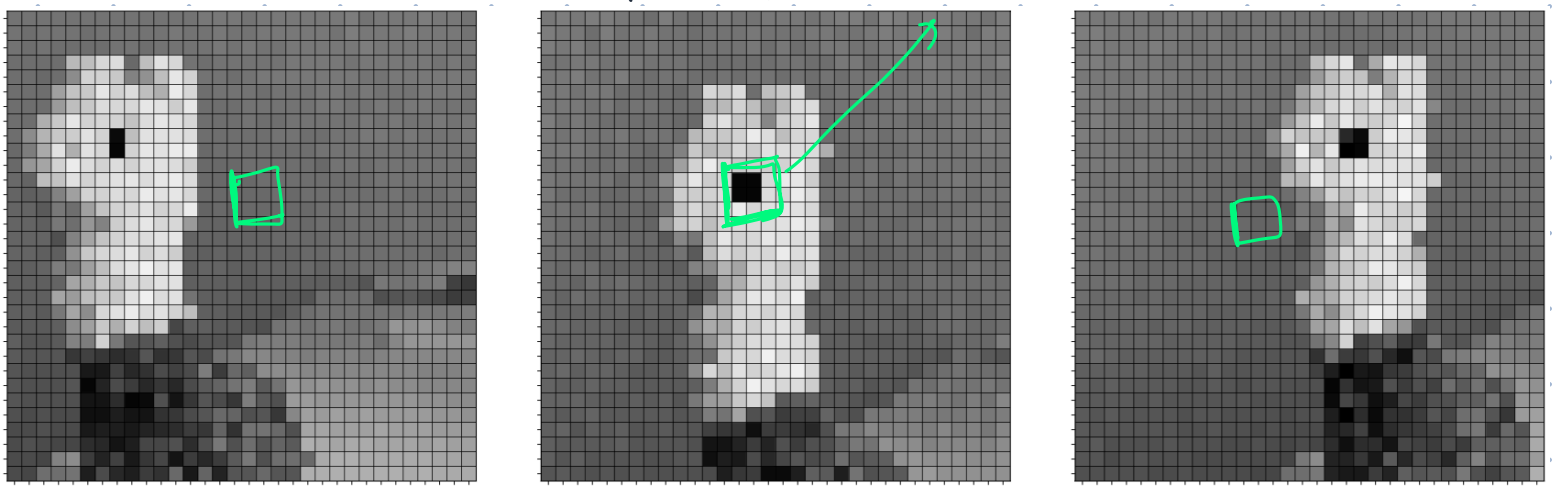


→ 300x300x3 (RGB) = 270,000 weights per neuron per layer! → Could overfit to training data!

but CS3980 students use 1024x1024x3

3 x 10^6 = 3M weights!

Beyond scalability – towards inductive bias



There might be more to life than a single pixel!

Goal: we need some architecture that is more tuned to image data

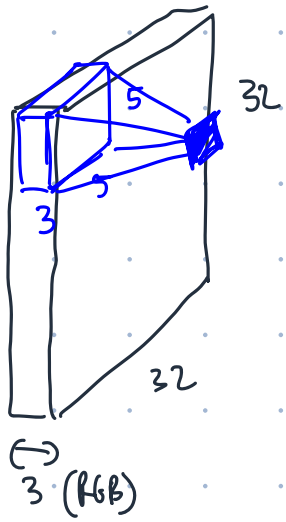
"inductive bias"

PRESERVING SPATIAL STRUCTURE

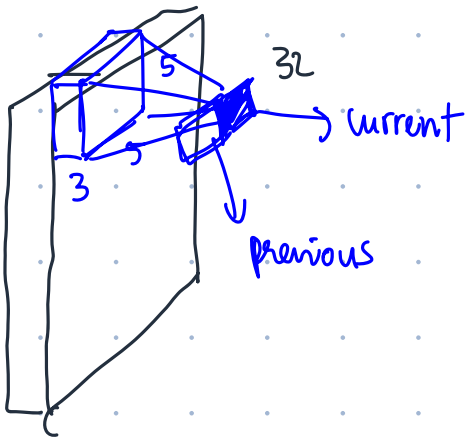
Pixel-by-pixel — BAD!

Idea — look at patches!

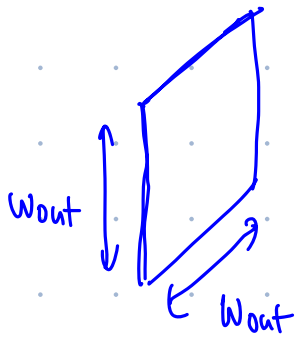
More specifically, "flop the filter, compute dot product, slide!"



Think: filter = detects the edge / neck of llama



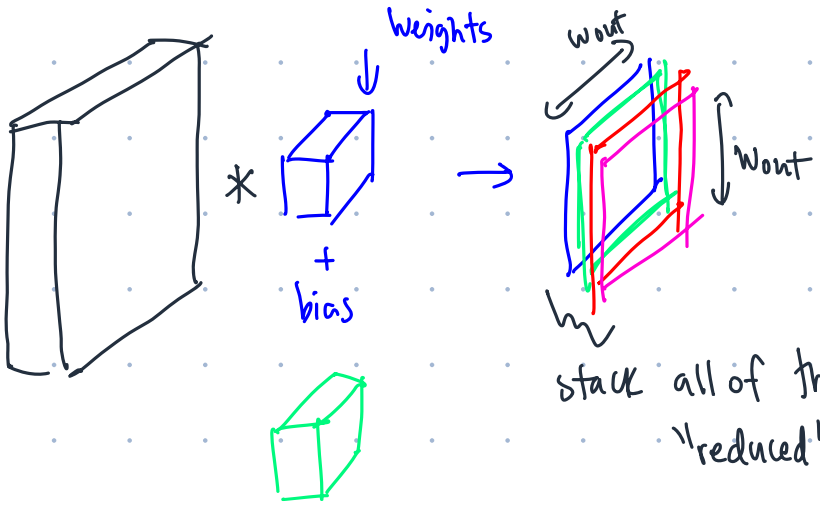
→ 2D



= output of convolution with that filter

More than one filter — Why not?

We can use more than one filter if we want



filters ≡ visual features

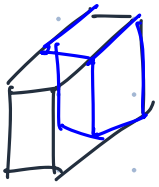
stack all of these to form a "reduced" volume

4 x Wout x Wout

Q. If we convolve a 32x32x3 image, using 5x5x3 filter, what would be the width/height of the output?

(Assume movement by one pixel every time!)

Aditya – 28x28x1



input img – 5x5x3, Conv w/ 5x5x3 filter = 1x1

5x6x3, " = 1x2

5x7x3, = 1x3

⋮

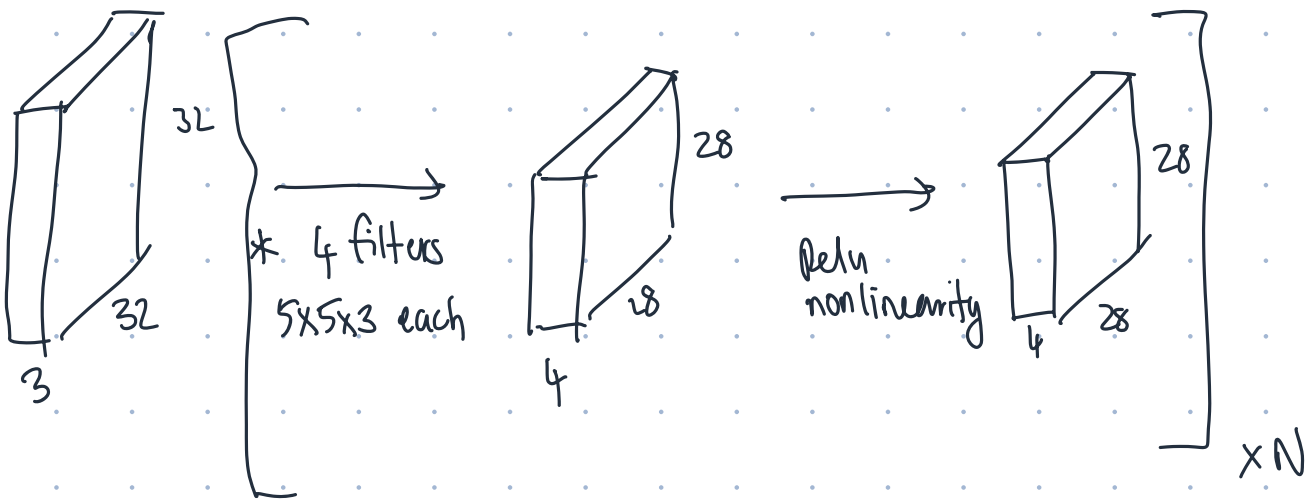
5x32x3 = 1x 32-5+1

= 1x 28

Final output would be 28x28

Convolution neural net = Conv + Relu

$W_2 W_1 x \equiv W_3 x$

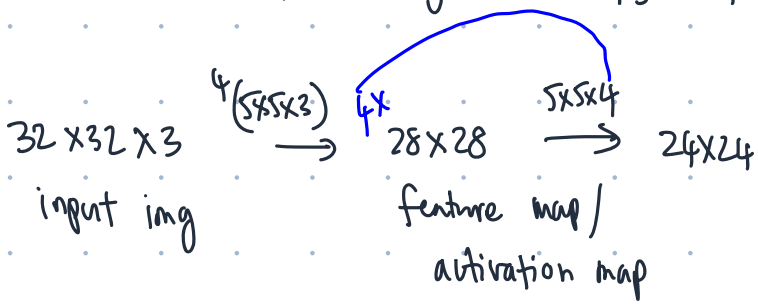


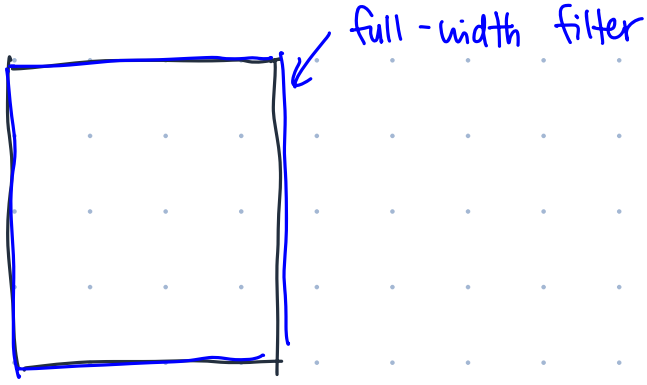
img → [conv+relu] * N → FCN (small) as the output neuron

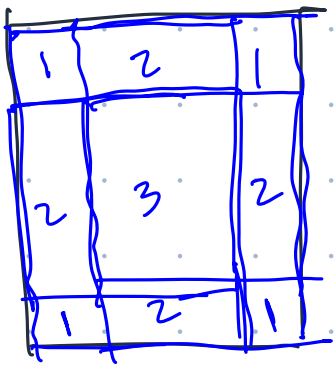
→ How do you train conv nets — backprop + SGD w/ momentum
w/ RMSprop

The Shrinkage problem —

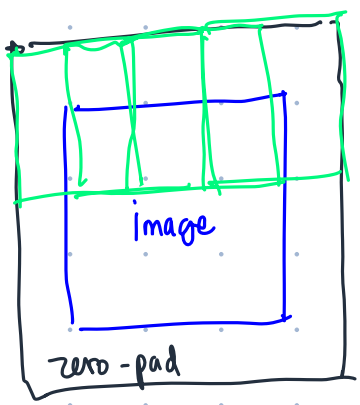
One convolution layer "shrinks" the output width







Every pixel is modeled equally



→ there is a configuration such that the output feature map has the same with as the input img (unpadded)

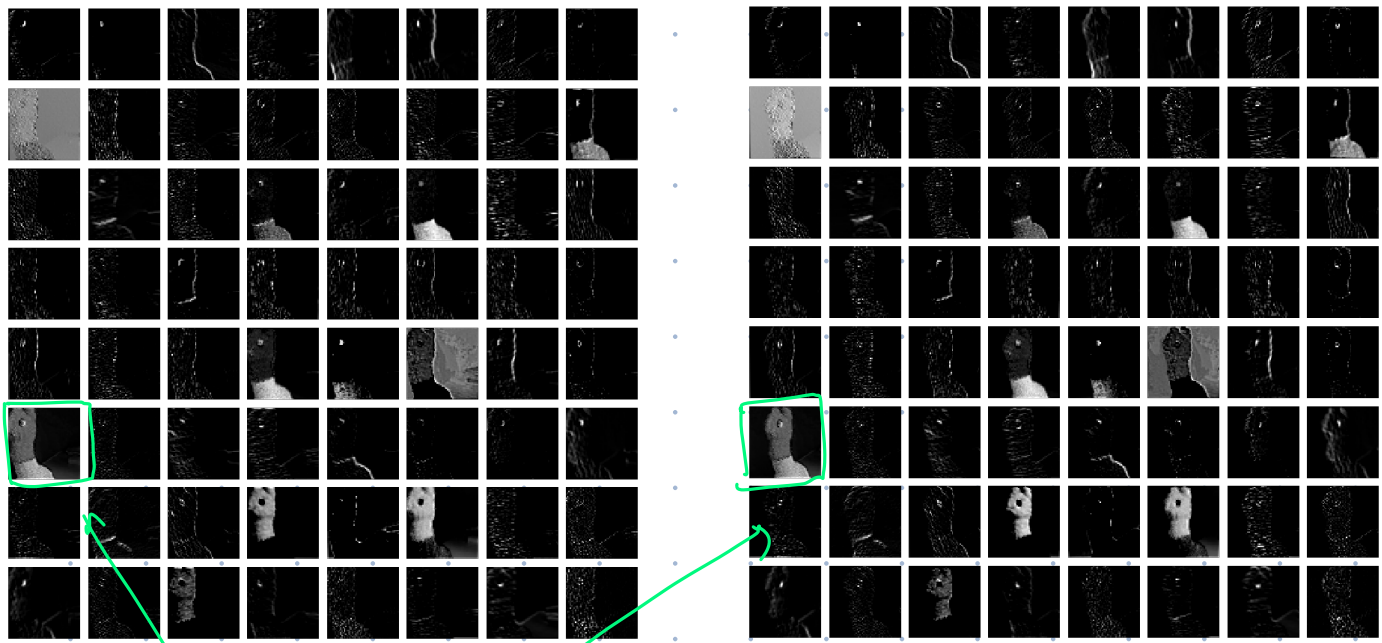
filter size

$$P = \frac{K-1}{2}$$

Translational equivariance –

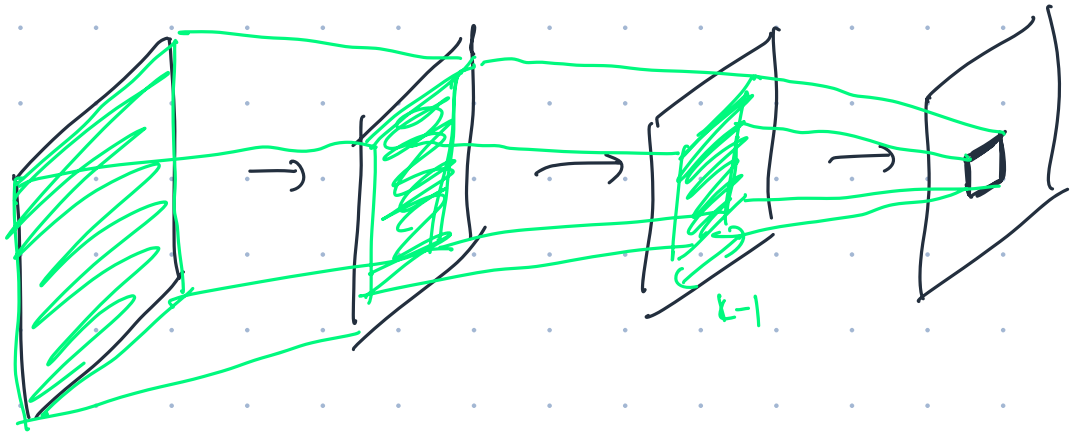
Conv gives you transl. Equivariance (if padding is right)

translate(conv(img)) ≡ conv(translated(img))



translated img, but feature detected!

Receptive fields (zero-padding)



Early layers capture simple patterns

later layers capture complex patterns

Need 'k' convolution layers to observe a pixel that "see's / encodes the whole image

If we were to use 1024 x 1024 imgs, we would need 500 / more conv layers to observe a single pixel that "sees" the whole img!

1. STRIDE

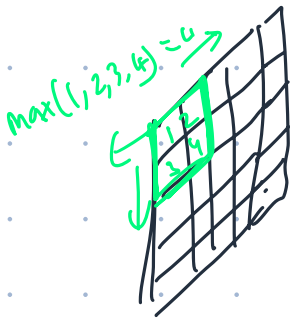
"Jump", don't move one pixel at a time, move 2/3/more

Simple idea: move in strides of two, say!

this increases the receptive area faster!!

2. POOLING

Do similar to convolution, but not convolution



Simply take the maximum value in my grid

≡ pooling layer, DOES NOT have any weights!

$$\text{img} \rightarrow \left[\left[\text{conv+relu}\right] \times N \rightarrow \text{pool}\right] \times M \rightarrow \text{FCN}$$

→ Avg pooling = blurred effect, smooths out outliers

→ Max pooling = would ensure sensitivity to minor shifts?

AlexNet!

Paper	Citation count (04/24/2025)
Darwin, "The Origin of The Species by Means of Natural Selection" (1859)	65,778
Shannon, "A Mathematical Theory of Communication" (1948)	66,335
Watson and Crick, "A Structure for Deoxyribose Nucleic Acid" (1953)	19,909
The ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC" (2012)	12,071
Krizhevsky, Sutskever, and Hinton, "ImageNet Classification with Deep Convolutional Neural Networks" (2012) [AlexNet]	142,510