CS 331
Computer Vision

11 / 16 / 2026

Instructor: Michael Eckmann

Topics

- Questions / Comments?
- More Convolutional Neural Networks
  - Receptive field
  - padding/stride
  - Pooling
  - How to determine the number of parameters and the sizes of outputs etc.

Convolutional Layer

- What's a tensor?
- What happens in a Convolutional Layer?

Convolutional Layer

- Suppose we have a batch of size N=32 color 128x128 images as input.
- Also suppose we are learning 64 kernels (aka masks, or filters) of size 3x3x3 in the convolutional layer.
- What are the dimensions of the output of that convolutional layer?
- How many weights/parameters are learned?

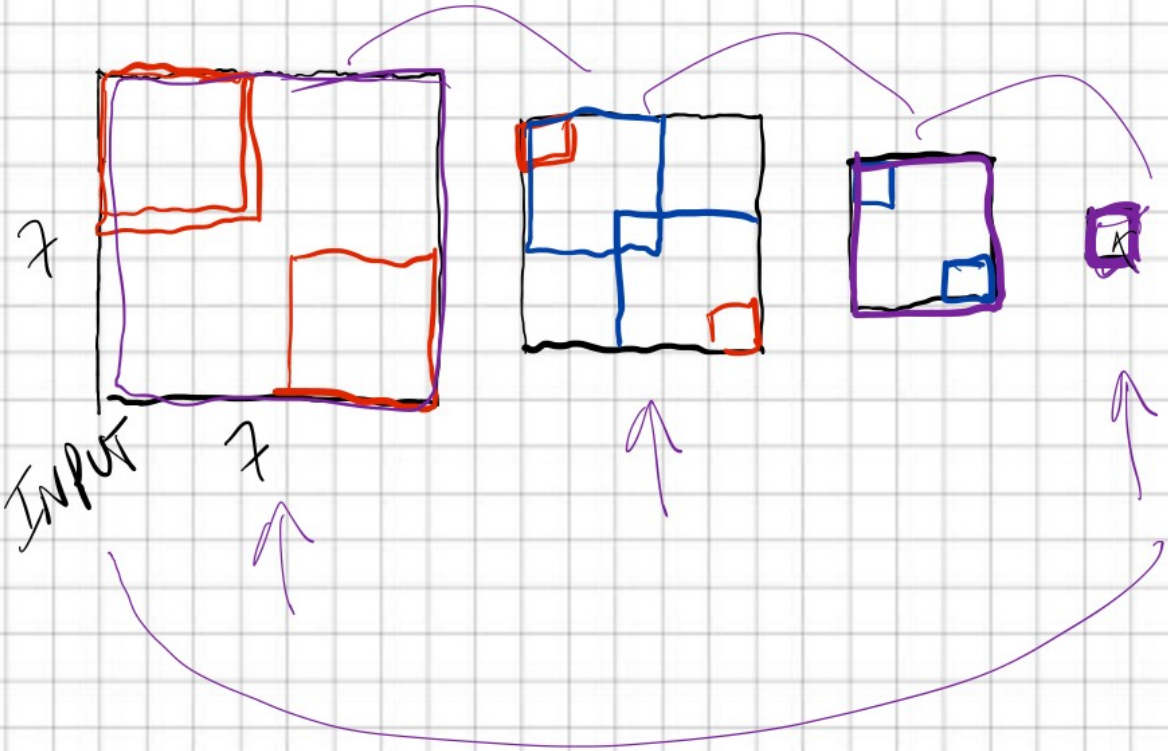
Convolutional Layer

- Suppose we have a batch of size N=32 color 128x128 images as input.
- Also suppose we are learning 64 kernels (aka masks, or filters) of size 3x3x3 in the convolutional layer.
- Assuming no padding and a stride of 1 the 128x128 reduces to 126x126 and 64 layers/channels also by N=32. What are the dimensions of the output of that convolutional layer? 32, 126, 126, 64
- How many weights/parameters are learned?
- All the weights are in the masks and the bias. Each mask has 9*9=27 weights plus 1 bias per mask. There are 64 of them so 64*(27 + 1) = 1792.

Convolutional Layer

- Receptive field --- the spatial portion of the input that affects the output of a convolutional layer
- Consider between 2 convolutional layers as well as from input layer all the way to the last convolutional layer
- Desirable for a later layer to have the entire input image as a receptive field --- to learn global spatial patterns
- The deeper the CNN the larger the receptive field.
- Downsampling also plays a significant role.

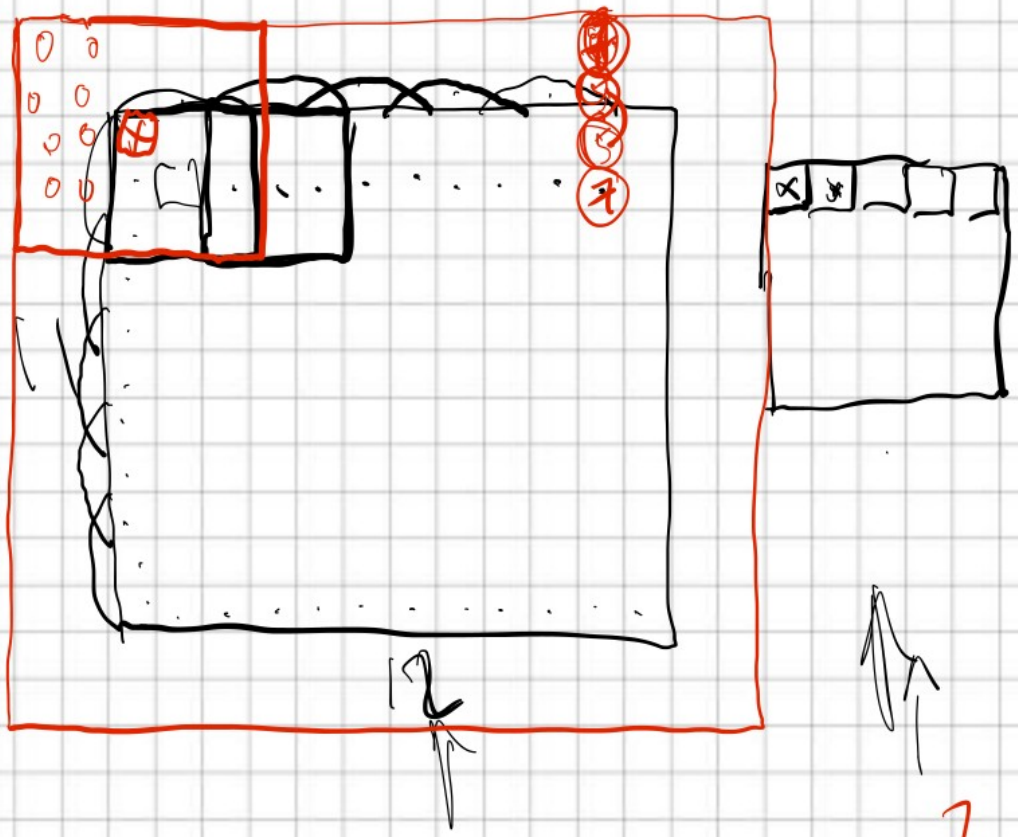
Receptive field



Convolutional Layer

- Could increase the stride to downsample, instead of making the network deeper.
- Stride is how much we move the mask after convolving.
- Until now, our stride was always 1.

Convolutional Layer (Stride)



PAD = 2
MASK = 5

PAD 1
MASK 3x3

Convolutional Layer

- Notice too how a convolution with a mask reduces the output spatial size (number of rows and number of columns)
- A solution to that is padding. We often pad the input image enough so that the output is same height and width. This is called “same”. Examples: pad 1 for a 3x3 mask, pad 2 for a 5x5 mask.
- Pad how much for a 7x7 mask for “same” padding?
- Zero padding refers to the value in the padded cells.
- Note: the padding size doesn’t always need to be same and the padding value doesn’t always need to be zero.

CNNs

- Downsampling can be done with a stride of 2 or more. A convolutional layer that does this convolution will involve masks and hence, learnable parameters.
- Another way to downsample, but without requiring any learnable parameters is a pooling layer, which also adds non-linearity.
- Hyperparameters
  - choice of pooling function (e.g. max or avg)
  - mask size
  - stride

Pooling Layer

MAX POOLING

MASK SIZE 2x2

STRIDE of 2

AVG

