Foundations of Convolutional Neural Networks

Sidharth Baskaran

September 2021

Computer Vision

- Image classification
  - Binary classification
- Object detection
  - Drawing boxes/bounding the objects
  - Multiple instances of object
- Neural Style Transfer
  - Content and style images
  - Repaint content w/ style
- Inputs can get large
  - E.g 64x64x3 is small but larger images have many input features

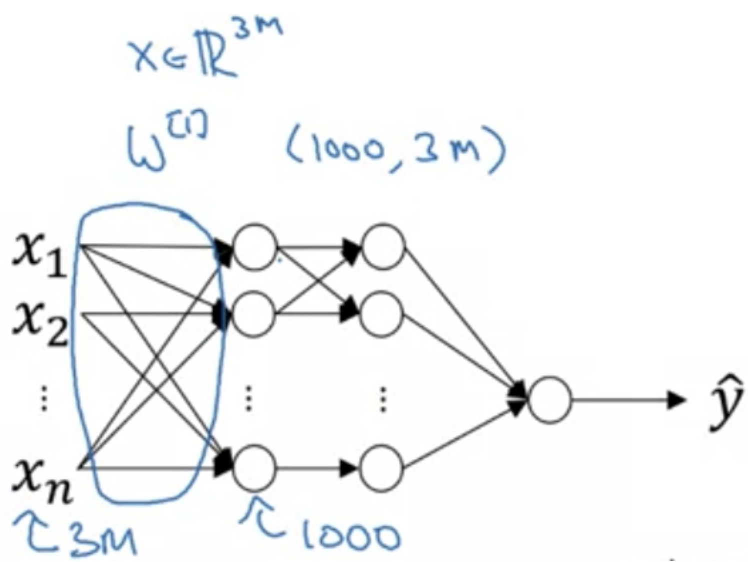

Figure 1: Example network

- If $x \in \mathbb{R}^{3M}$ then $W^{[1]} \rightarrow (1000, 3M)$
  - $W^{[1]}x$ gives the output network vector of dimension $(1000, 1)$
- Implementing the convolution operation is efficient for large input images

Edge Detection

- Detecting certain festure sets of images
  - E.g. vertical and horizontal edges

- Given a 6x6 grayscale image, apply a 3x3 kernel or filter
  - **Convolution operator** $*$ convolves filter over image
- Paste filter on the first such region of the image, and take elementwise product and then sum to obtain value
- Output a 4x4 image

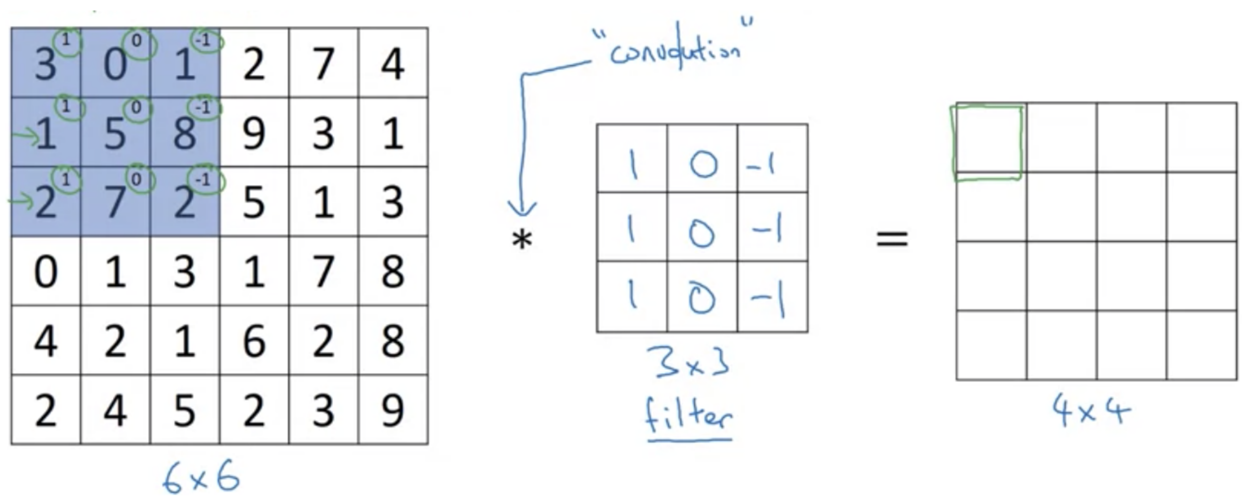

Figure 2: Basic Convolution

- Shift the kernel stepwise to the left to fill up the output
- Is 4x4 as can shift downwards/left/right to obtain 4 unique locations
  - $\dim(\mathrm{Im}, 1) - \dim(\mathrm{Ker}, 1) + 1$

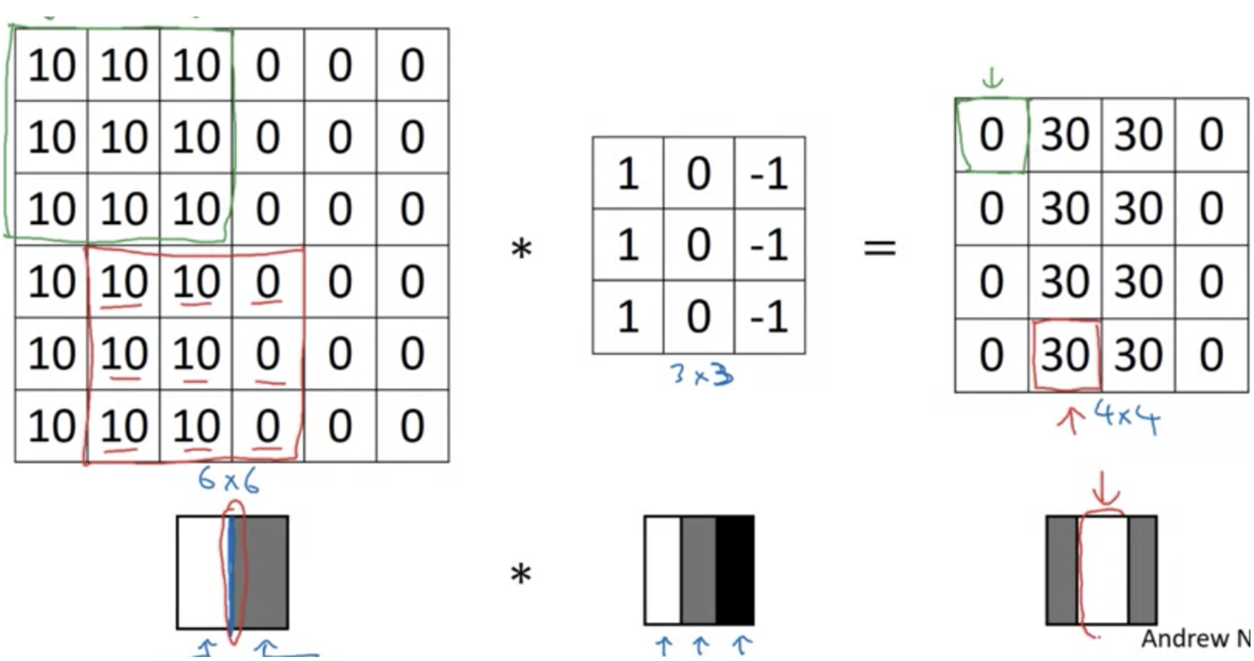

Figure 3: Example

- In example, detects a light to dark transition
- Can also make distinction

- *Example*: Sobel filter $\rightarrow$ puts weight towards central pixel, more robust for edge detection

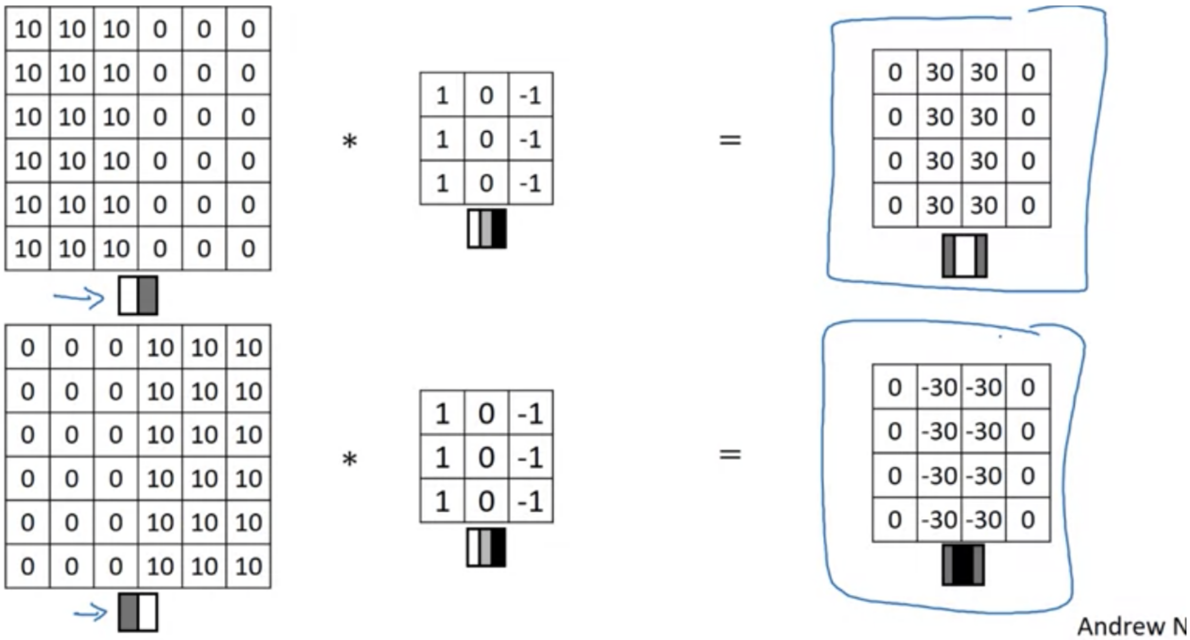

Figure 4: Transition examples

- Rotate 90 degrees for horizontal/vertical differentiation
- Can learn the filter values with backprop
  - Treat as parameters, learn

Padding

- Given an image with input dimensions $n \times n$ and filter with $f \times f$
  - Output dimensions are $n - f + 1 \times n - f + 1$
- Downsides $\rightarrow$ image shrinks
  - Lots of overlap on central pixels but corner pixels represented less
- Can pad image with border of 1px, for example
  - 6x6 $\rightarrow$ 8x8 image
- Can preserve dimension of output image
  - Padding pixels are 0 by convention
  - Let $p = 1$ be padding amount
- Valid and Same convolutions
  - Valid $\rightarrow$ no padding
    - $n \times n \ * \ f \times f \rightarrow n - f + 1 \times n - f + 1$
  - Same $\rightarrow$ pad s.t. output size = input size
- By convention, f is always odd, i.e $f \mod 2 \neq 0$

Strided Convolutions

- Stride of s means moving kernel by s steps instead of default 1
- Output dimensions become following where p is the padding amount and s is the stride

$$\left\lfloor \frac{n + 2p - f}{s} + 1 \right\rfloor \times \left\lfloor \frac{n + 2p - f}{s} + 1 \right\rfloor$$

- Floor because do not want kernel being outside of the image or padding region

- A preprocessing step on convolution in mathematics
  - Flip the kernel vertically then horizontally (mirroring)
- Use flipped kernel for convolution $\rightarrow$ cross-correlation
  - Not required in NNs
- Convolution obeys property of being associative but not commutative
  - $(A * B) * C = A * (B * C)$