EECS 182/282A

Today: convolutional neural nets

Reading: Prince through Ch 9 (so far)
Ch 10 & 11 now.

Architecture Order In Class:

MLPs $\rightarrow$ CNNs $\rightarrow$ Graph NN $\rightarrow$ RNN/State-space $\rightarrow$ Transformers

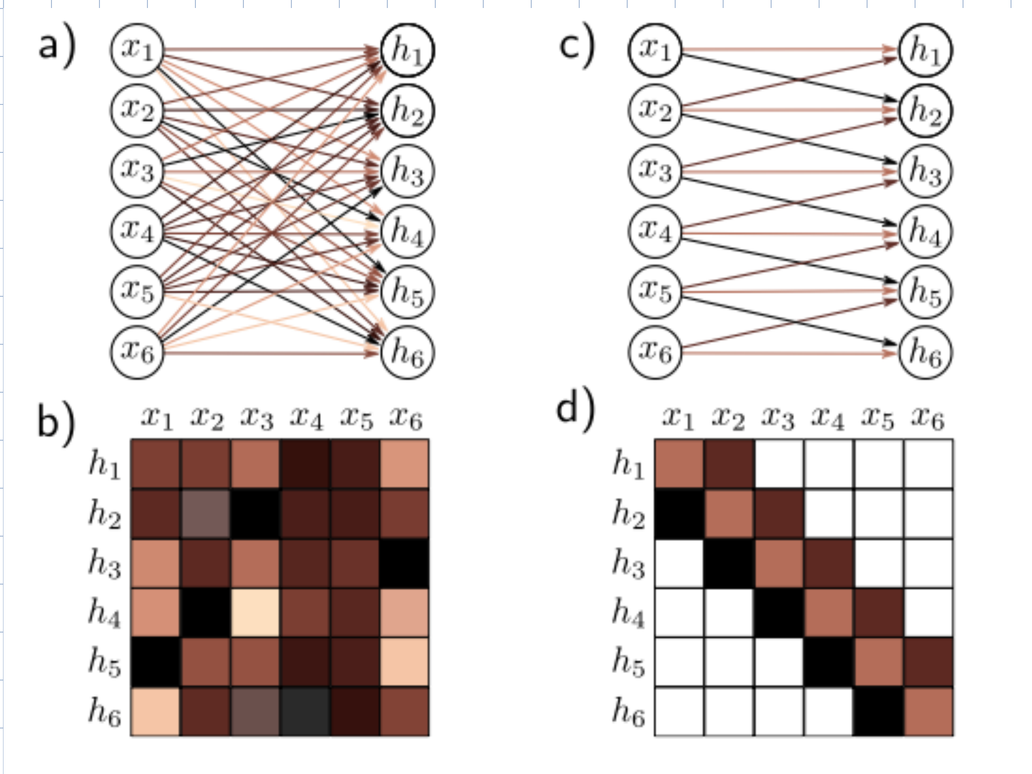


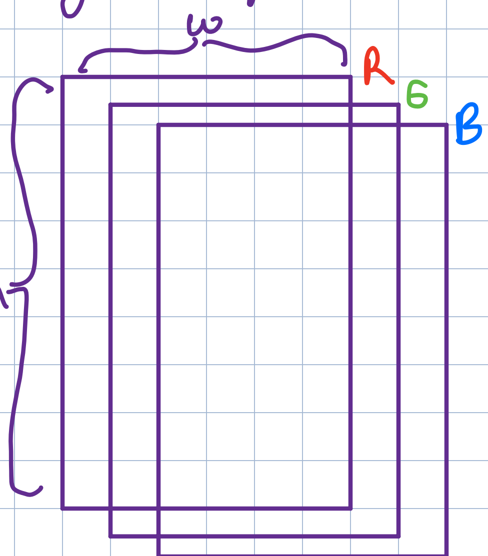
Fig 10.4 in Prince

MLP

CNN

Inspired by computer vision problems: classification,

Images



$h \times w \times 3$

channels c

semantic segmentation, etc.

Key Ideas (Expressivity)

1) Respect locality $\longrightarrow$ convolutional structure with small filters.

2) Respect certain invariances / equivariances / symmetries → weight sharing
   → data augmentations

   training

3) Hierarchical Structure (Parts make wholes) → Depth
   → Multiresolution & Filterbanks

Key Ideas (Getting it to work)

1) Normalization Layers
2) Dropout
3) Residual / Skip Connections

Recall 1D convolution (LTI systems... e.g. from EECS 120, 123)

$$y(t) = \int_{-\infty}^{+\infty} x(\tau)\, h(t-\tau)\, d\tau = \int_{-\infty}^{+\infty} h(\tau)\, x(t-\tau)\, d\tau$$

$$y[t] = \sum_{\tau=-\infty}^{+\infty} x[\tau]\, h[t-\tau] = \sum_{\tau=-\infty}^{+\infty} h[\tau]\, x[t-\tau] \xRightarrow{\text{FIR}} \sum_{\tau=-k}^{+k} h[\tau]\, x[t-\tau]$$

In Deep Learning, we don't "flip"

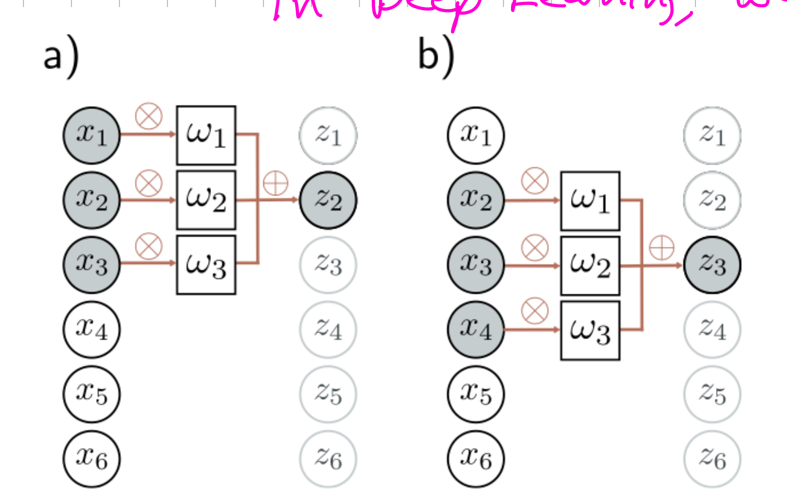


Figure 10.7 in Prince.

Notice: No flip

Can also have a bias

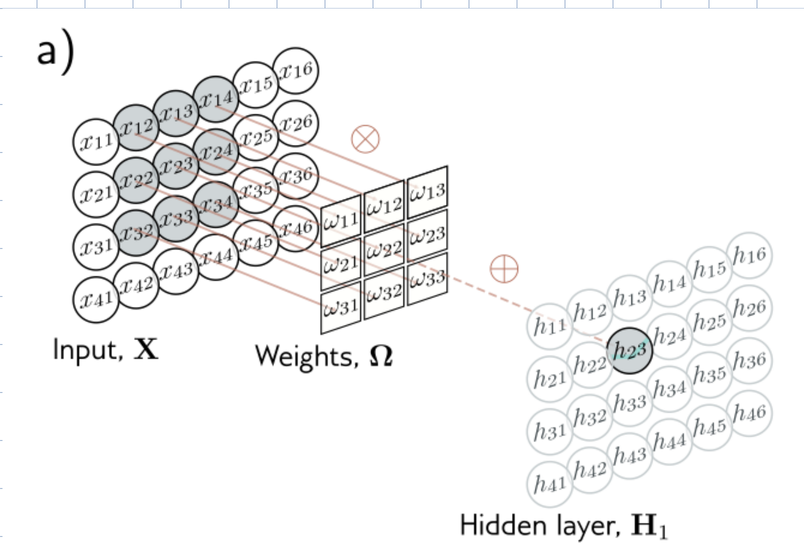


Figure 10.9 in Prince

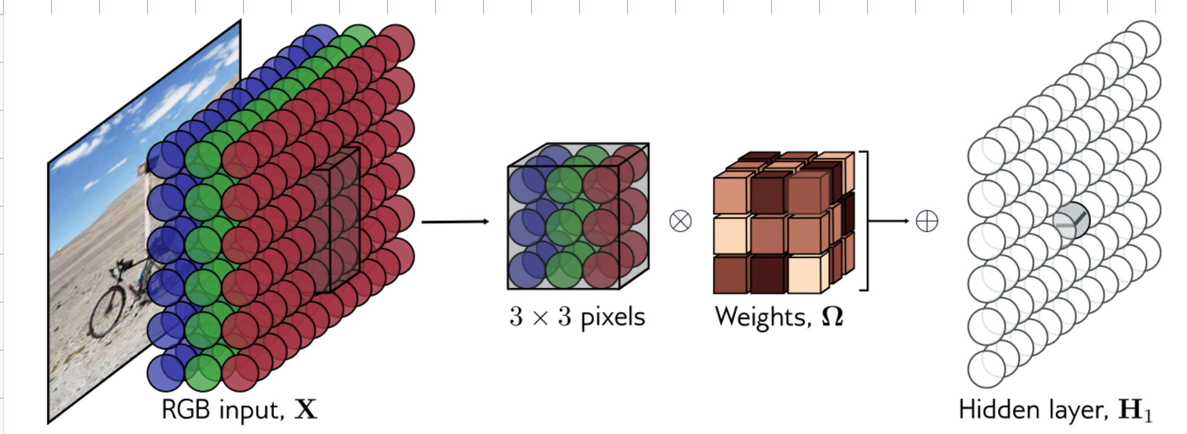
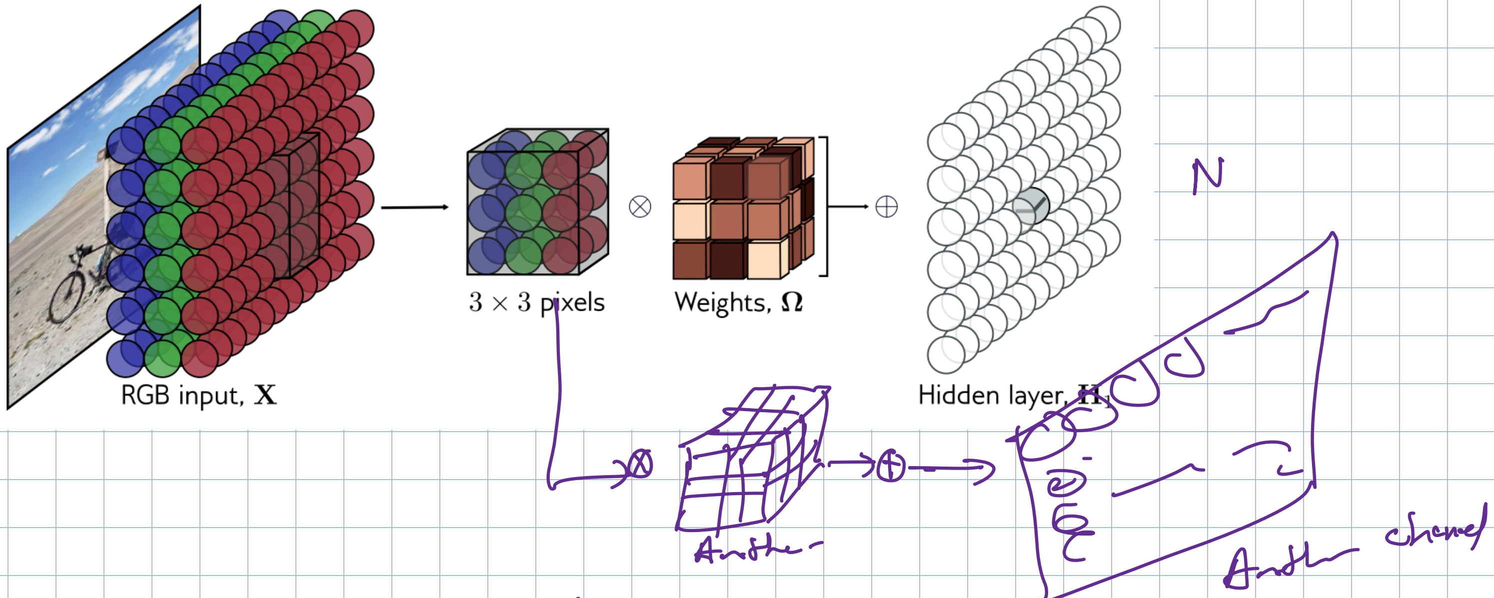


Figure 10.10 in Prince

Note: weights NOT shared across channels.



If we want more channels at the output of the layer, can repeat with fresh weights & biases for each output channel.

Can think (inaccurately) about output channels at layer ℓ as being X-detectors for that spatial position and different Xs.
e.g., fingers, toes, flies, ... for X.

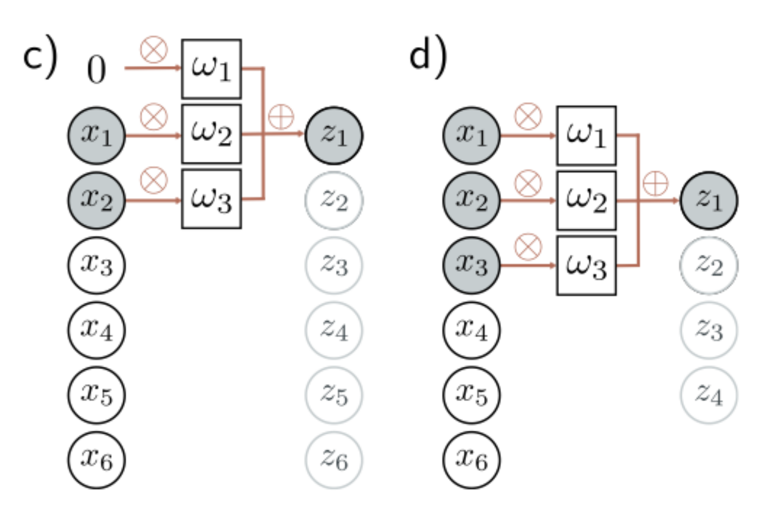
<u>Special Case</u>: 1×1 conv with c_{in} input channels and c_{out} output channels

Input (from one pixel) is a vector $\vec{x} \}\, c_{in}$

Output (for this pixel) is a vector $\vec{y} \}\, c_{out}$

$$\vec{y} = \underbrace{\overbrace{W}^{c_{in}}}_{c_{out}} \vec{x} + \vec{b} \}\, c_{out}$$

Can also have a nonlinearity. e.g. $\vec{y} = \mathrm{ReLU}(W\vec{x} + \vec{b})$



"Same"-Padding　　No Padding

Figure 10.2 in Prince

Padding: Choices?

0-padding

Wrap-around padding
(circular)

Copy - padding

Avg -

Mirror-padding.

Receptive Fields

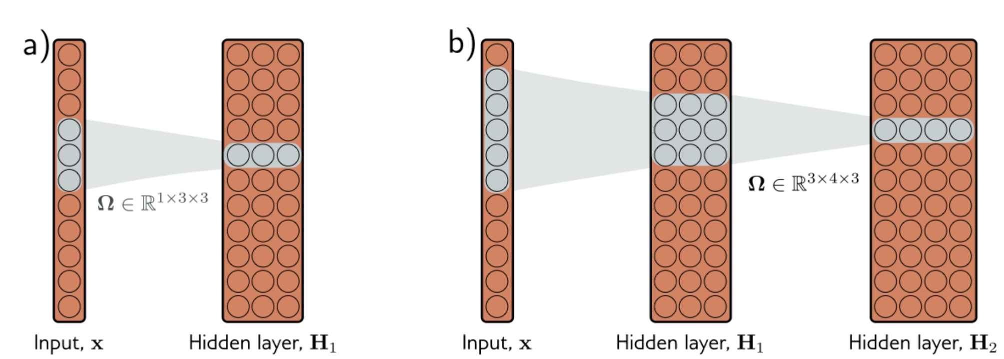


Figure 10.6 in Prince

Suppose all convs are 3x3
(or in 1d 3)

At layer ℓ, how big is receptive field?

Ans: $2\ell+1$

Downsampling via stride

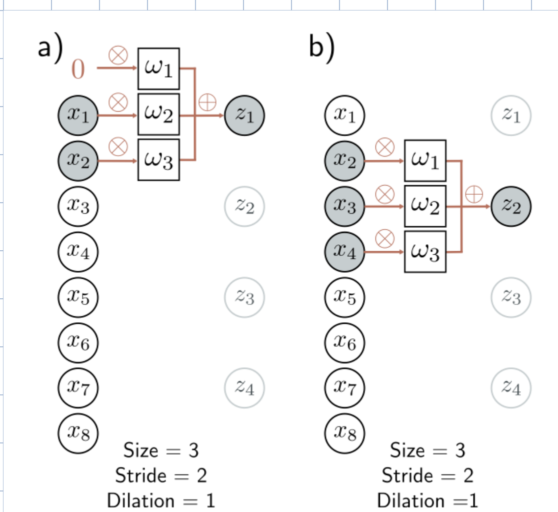


Figure 10.3 in Prince

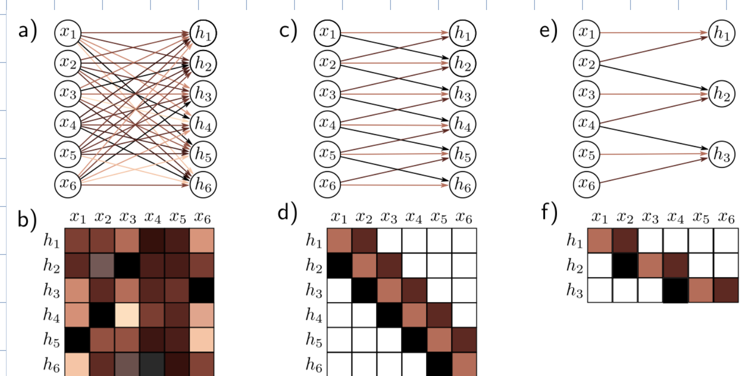


Fig 10.4

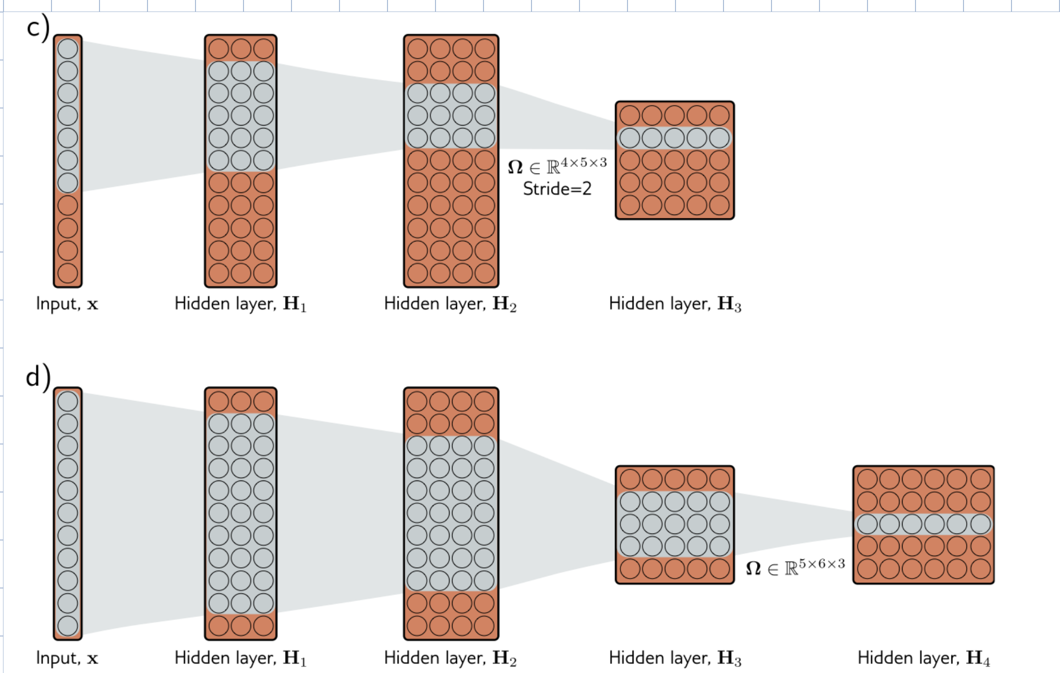


Figure 10.6 in Prince

Pooling

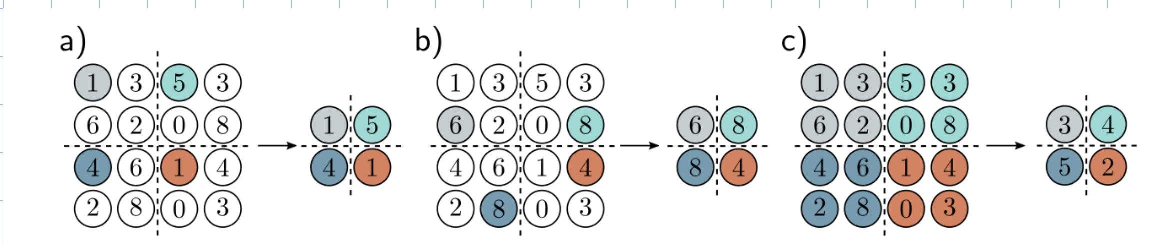
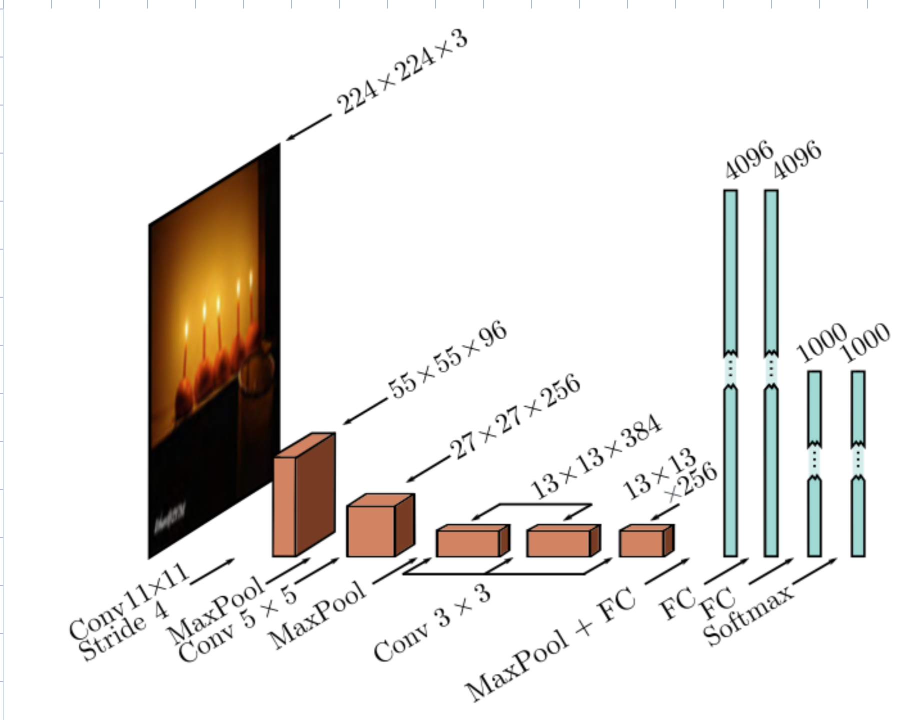


Fig 10.11 in Prince

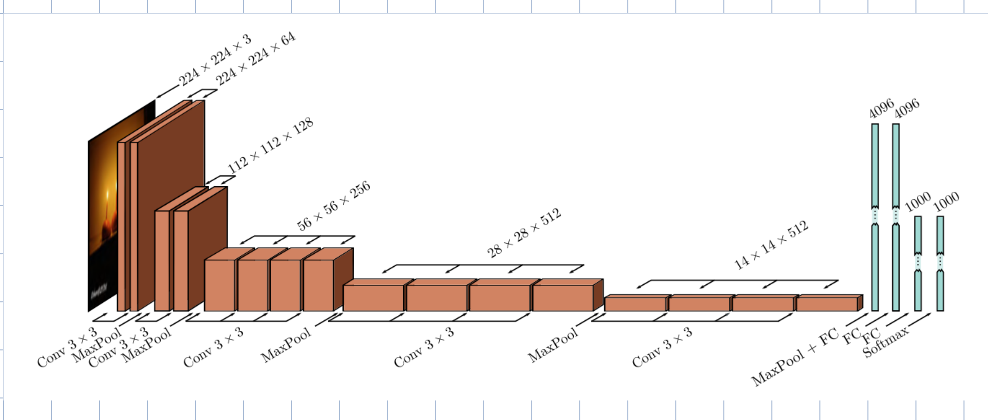
stride/subsample　　Max-Pool　　Mean-Pool

Putting things together



AlexNet (Krizhevsky et al 2012)

Fig 10.16 in Prince



VGG from 2014

Fig 10.17 in Prince