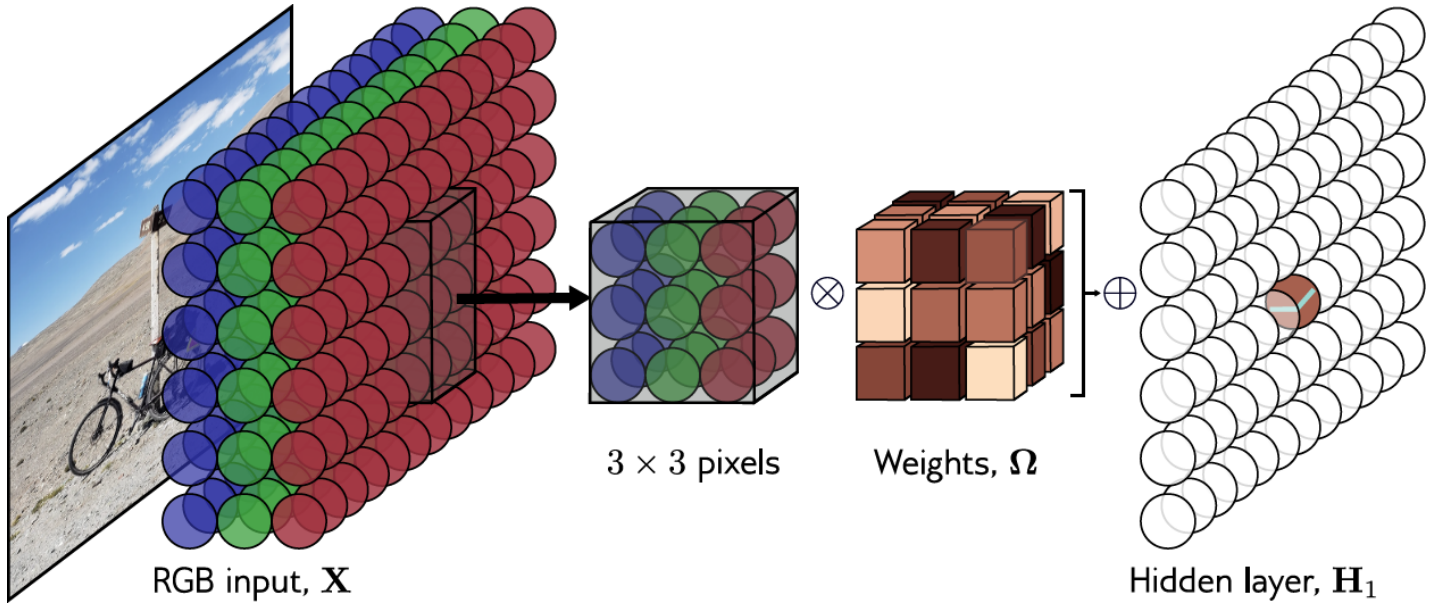
Advanced Deep Learning

DATA.ML.230

Reminder: 2D CNNs

From: Simon J. D. Prince, Chapter 10 – Understanding Deep Learning, MIT Press (19 May 2025)

2D Convolution and 2D CNNs

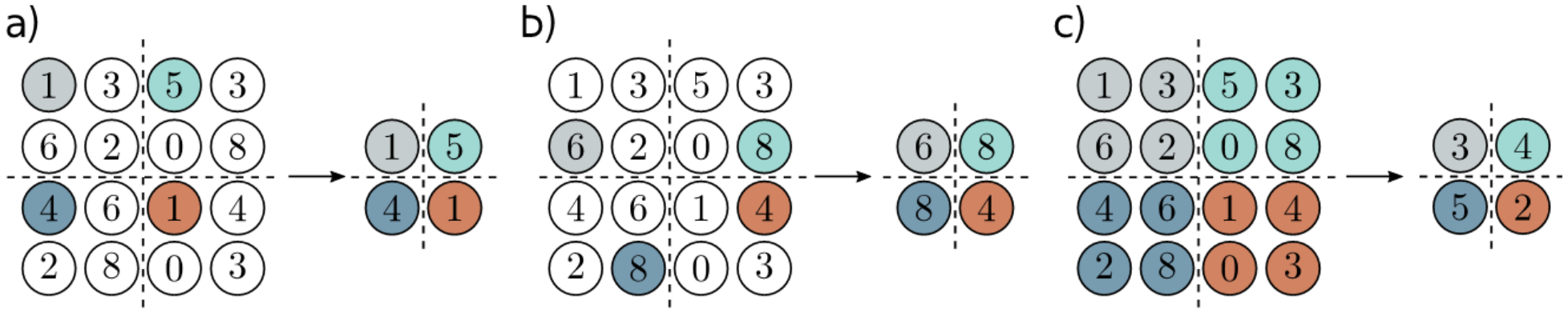


- Convolution in 2D
  - Weighted sum over a K x K region
  - K x K weights
- Build into a convolutional layer by adding bias and passing through activation function

$$h_{i,j} = \mathrm{a}\left[\beta + \sum_{m=1}^{3}\sum_{n=1}^{3} \omega_{m,n} x_{i+m-2,j+n-2}\right]$$

2D Convolution and 2D CNNs

Downsampling



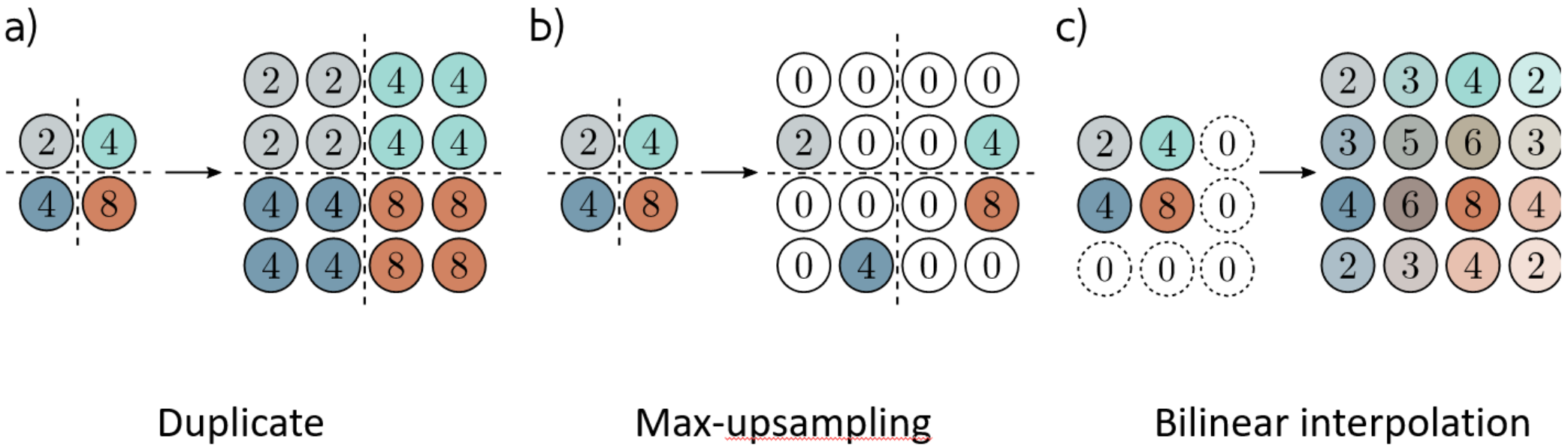
Sample every other position (equivalent to stride two)

Max pooling (partial invariance to translation)

Mean pooling

2D Convolution and 2D CNNs

Upsampling



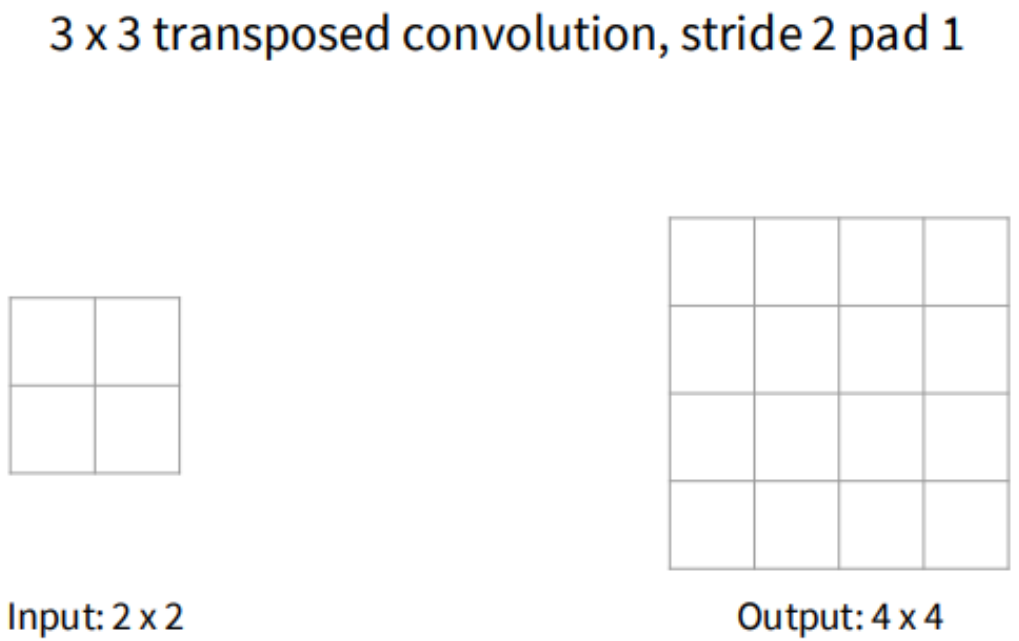
Duplicate

Max-upsampling

Bilinear interpolation

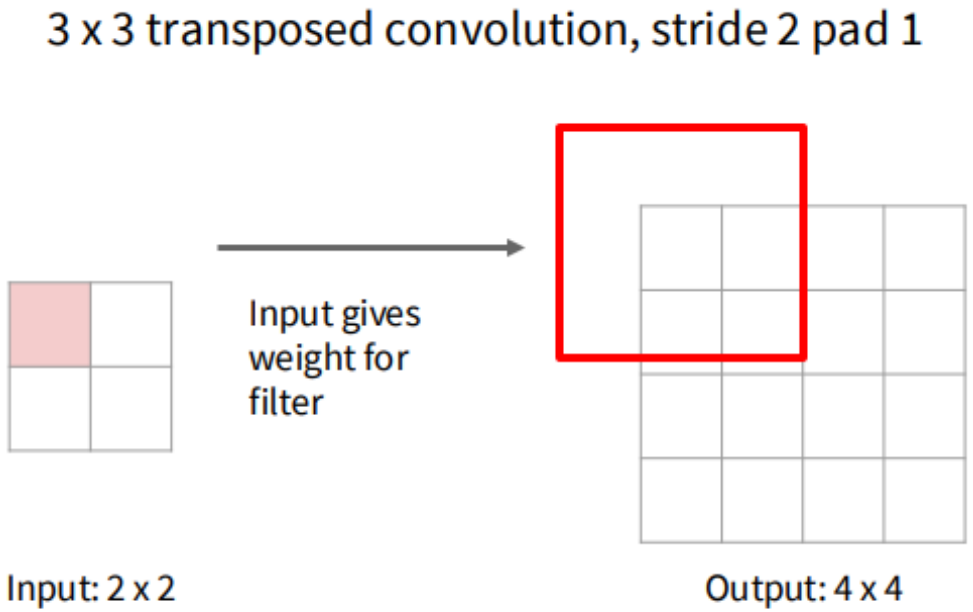
2D Convolution and 2D CNNs

Learnable upsampling



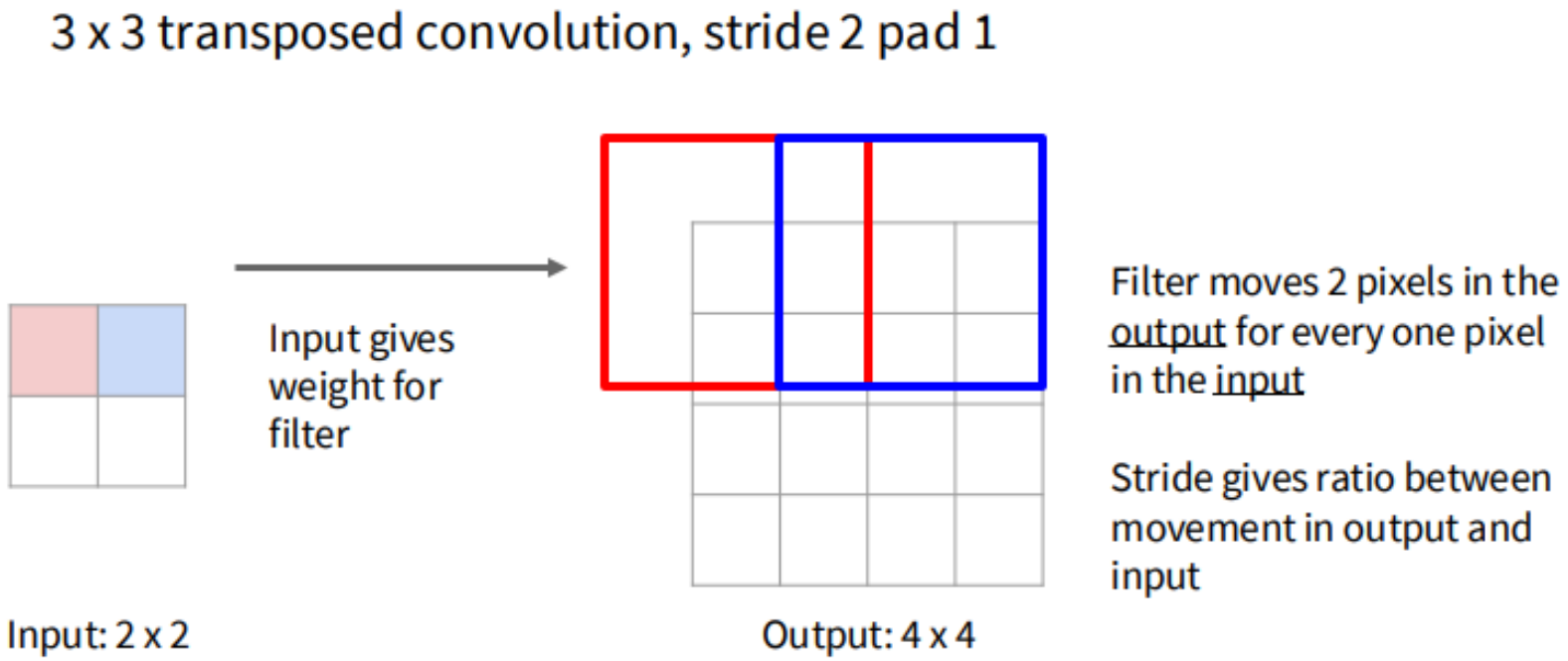
2D Convolution and 2D CNNs

Learnable upsampling



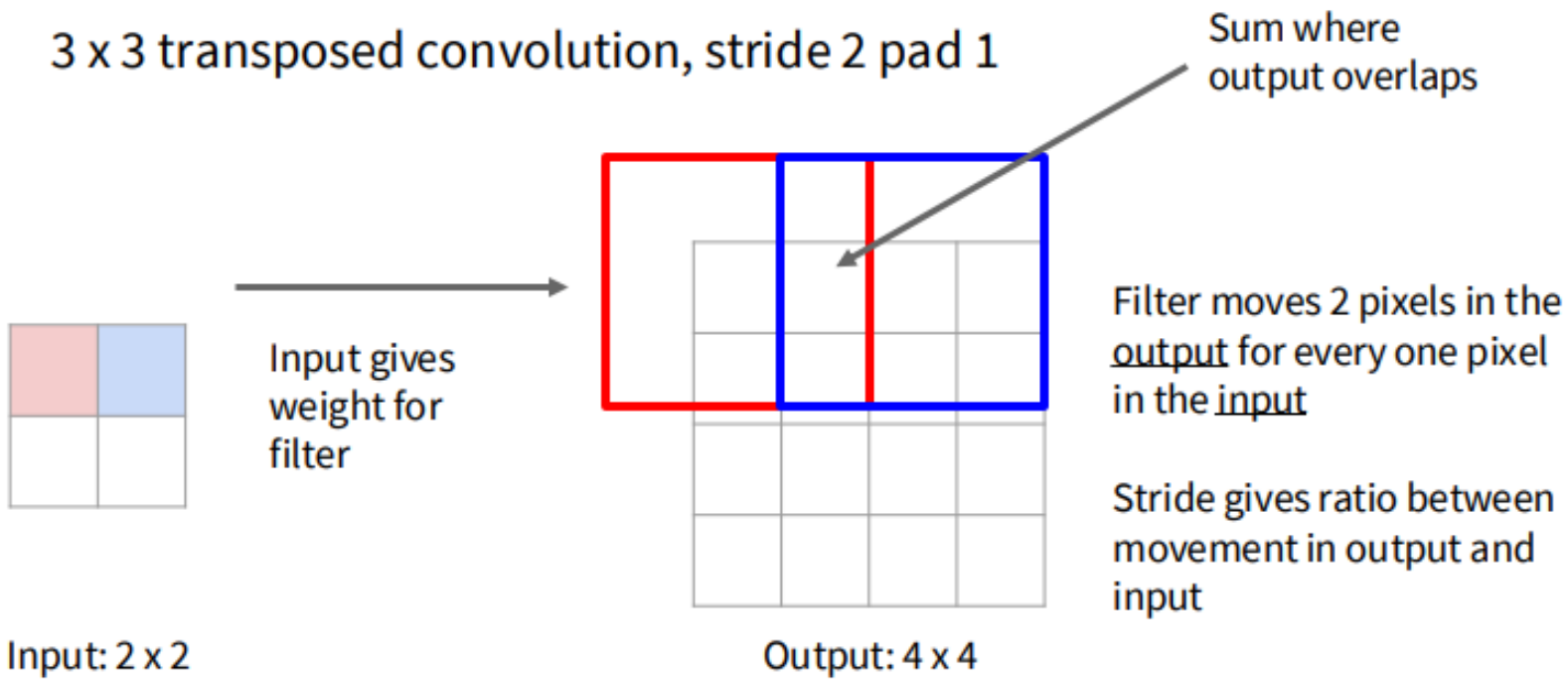
2D Convolution and 2D CNNs

Learnable upsampling



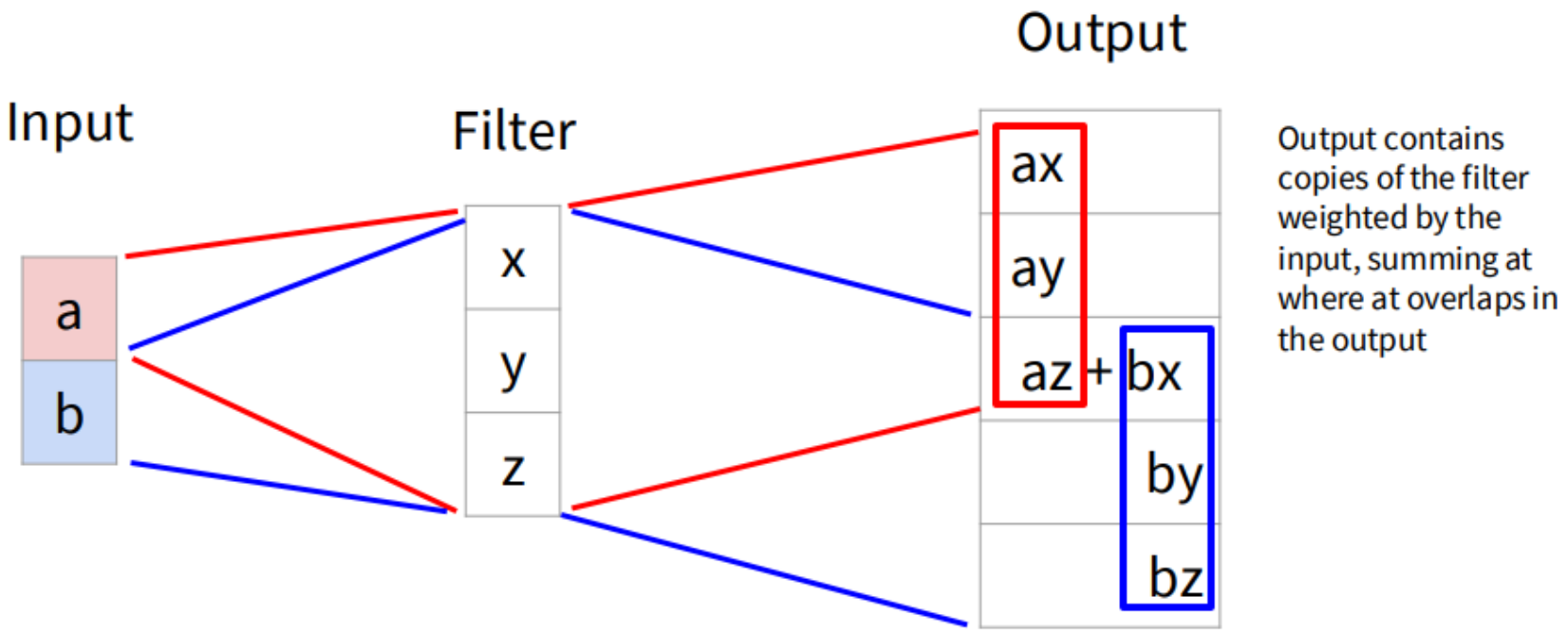
2D Convolution and 2D CNNs

Learnable upsampling



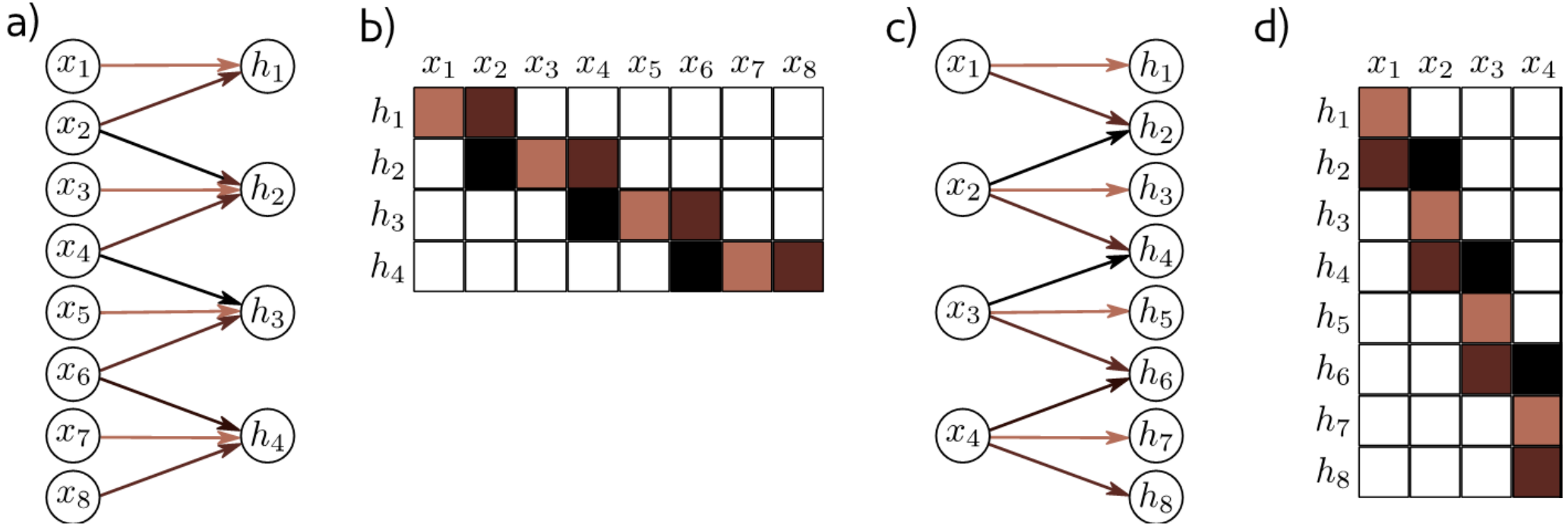
2D Convolution and 2D CNNs

Learnable upsampling: 1D example



2D Convolution and 2D CNNs

Transposed convolutions

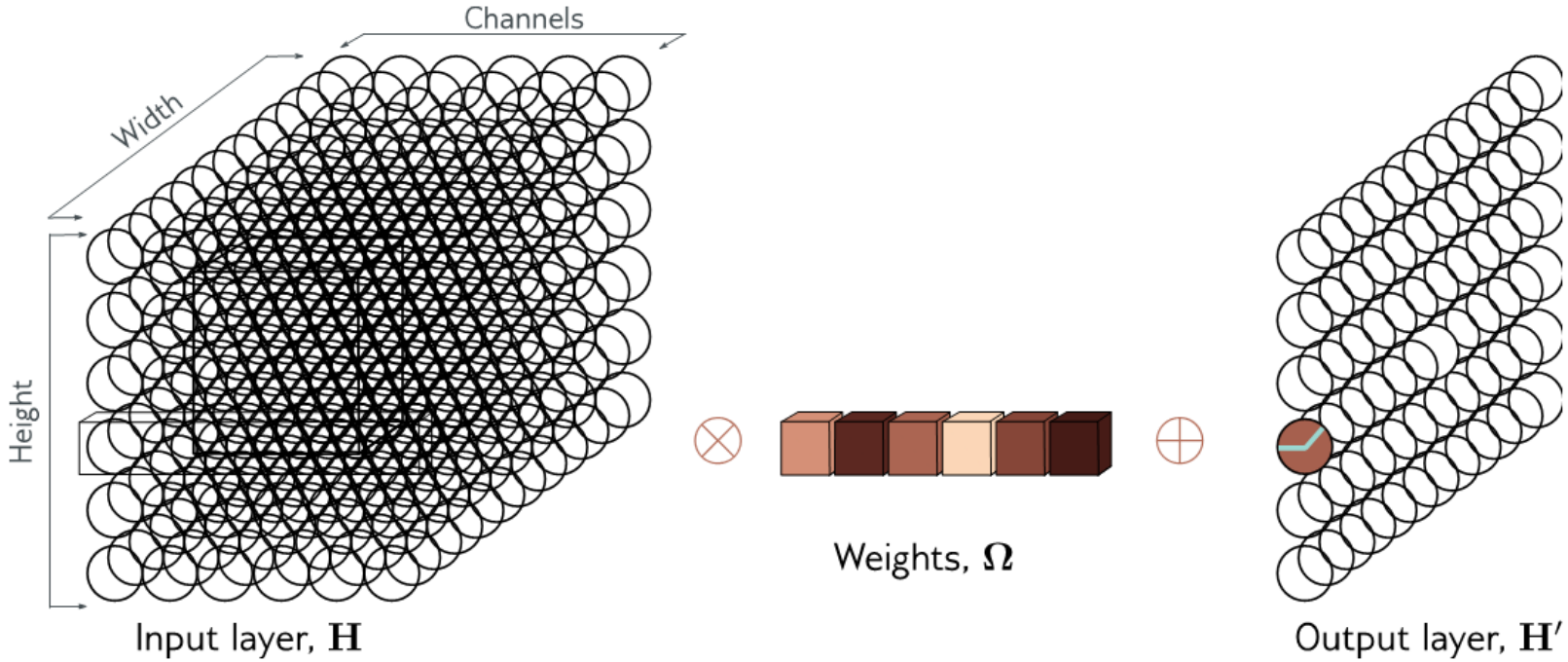


Kernel size 3, Stride 2 convolution

Transposed convolution

2D Convolution and 2D CNNs

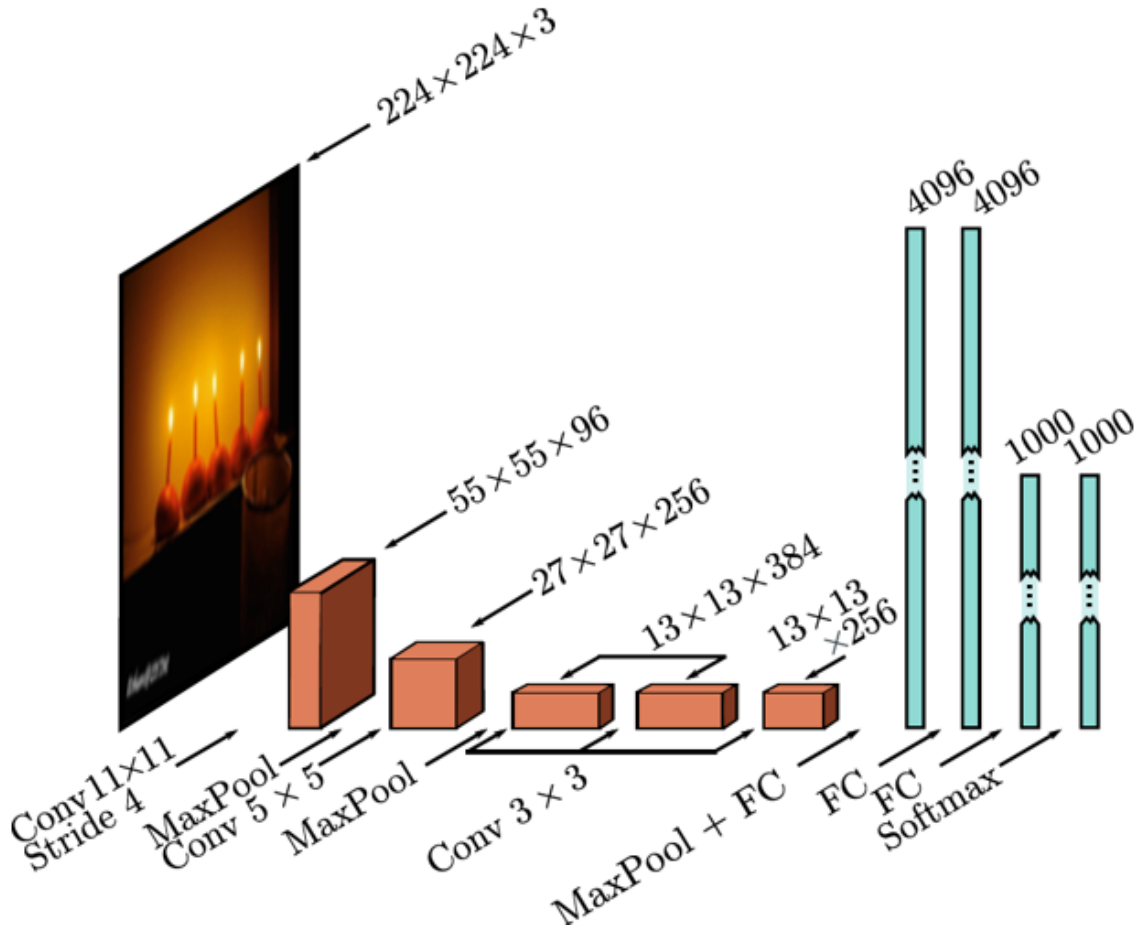
1x1 convolution



- Mixes channels
- Can change number of channels
- Equivalent to running same fully connected network at each position

2D Convolution and 2D CNNs

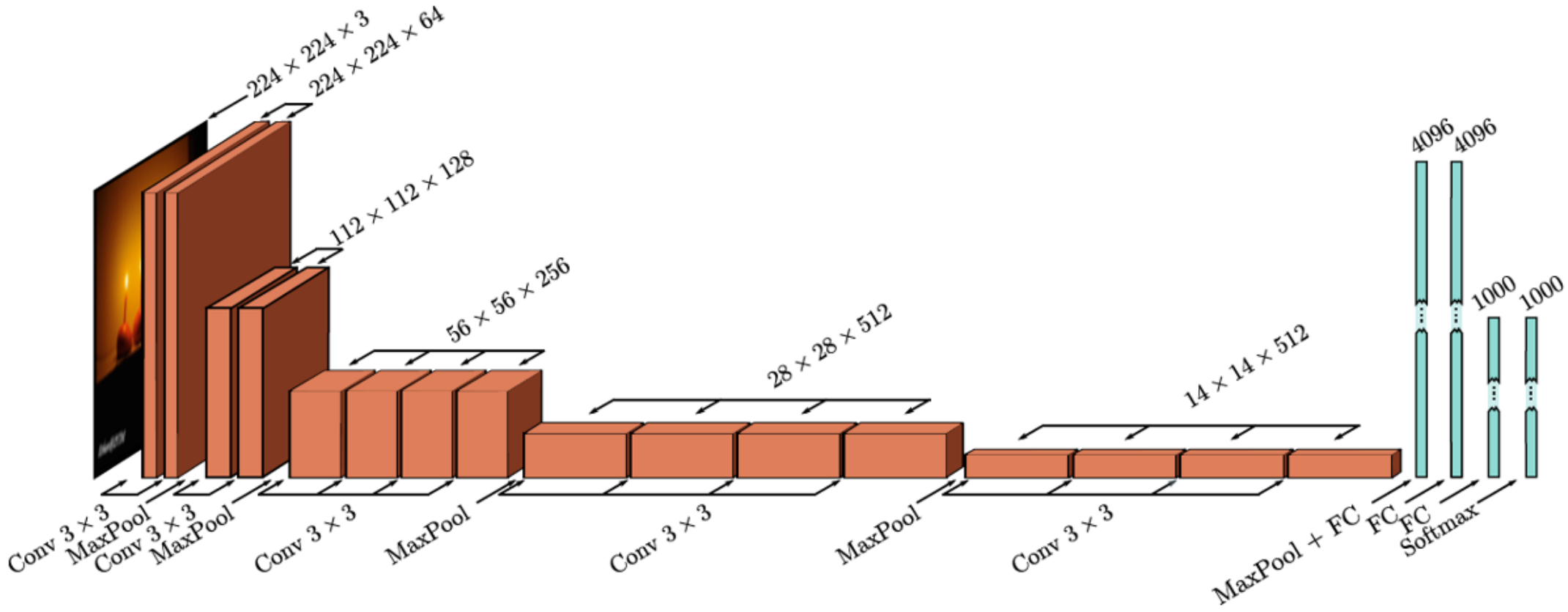
AlexNet (2012)



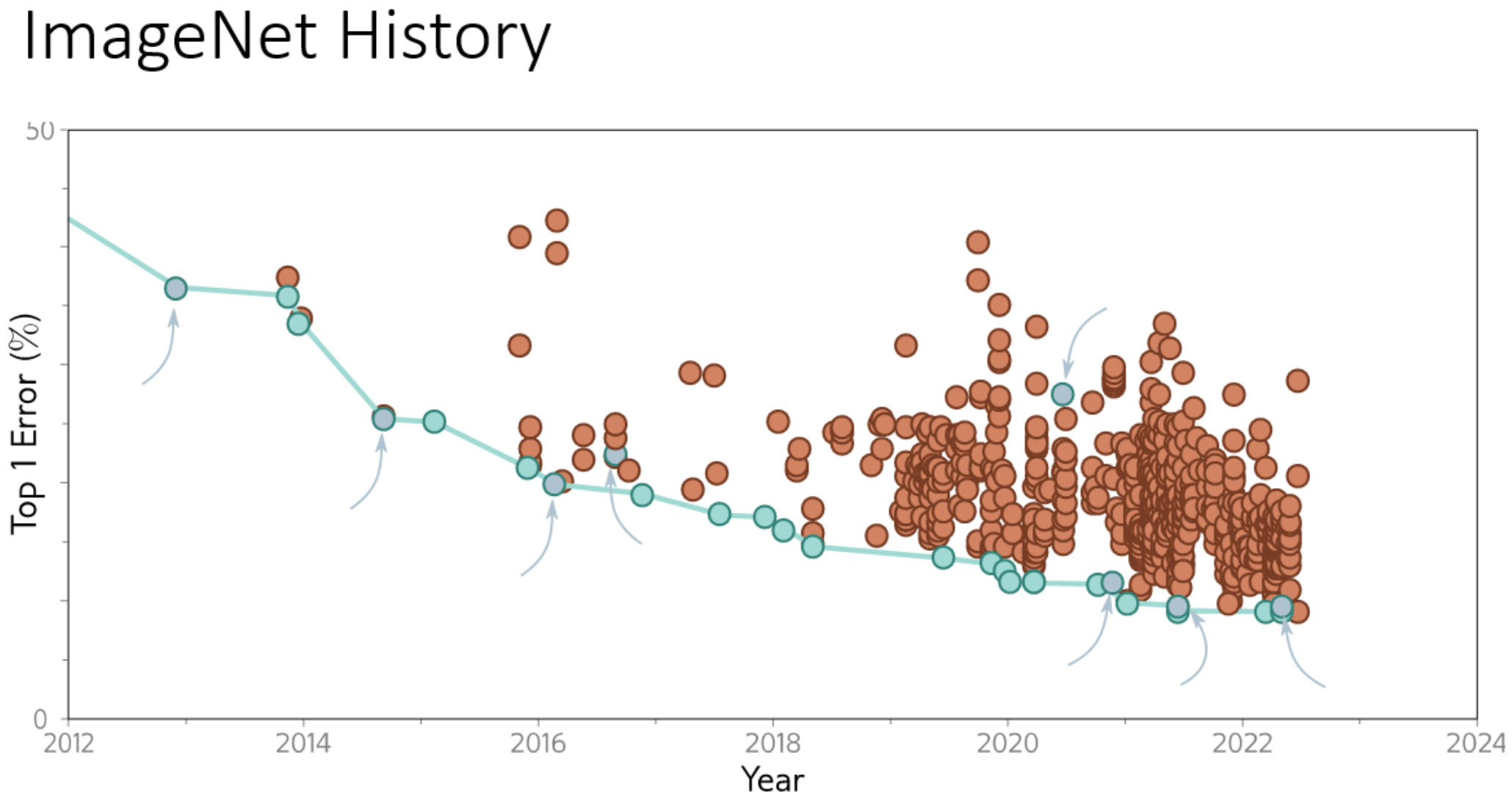
Almost all the 60 million parameters
parameters are in fully connected layers

2D Convolution and 2D CNNs

VGG (2015)

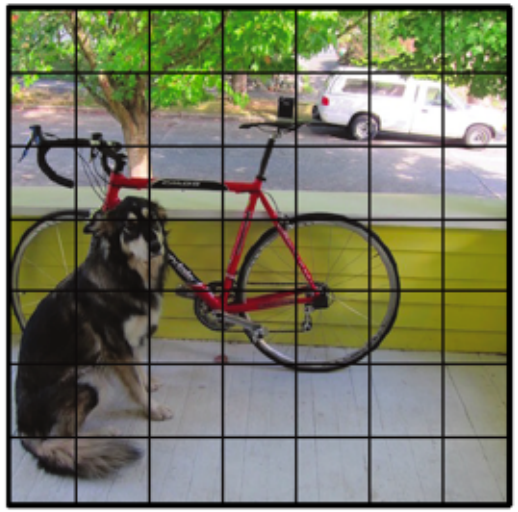


2D Convolution and 2D CNNs



2D CNNs for object detection

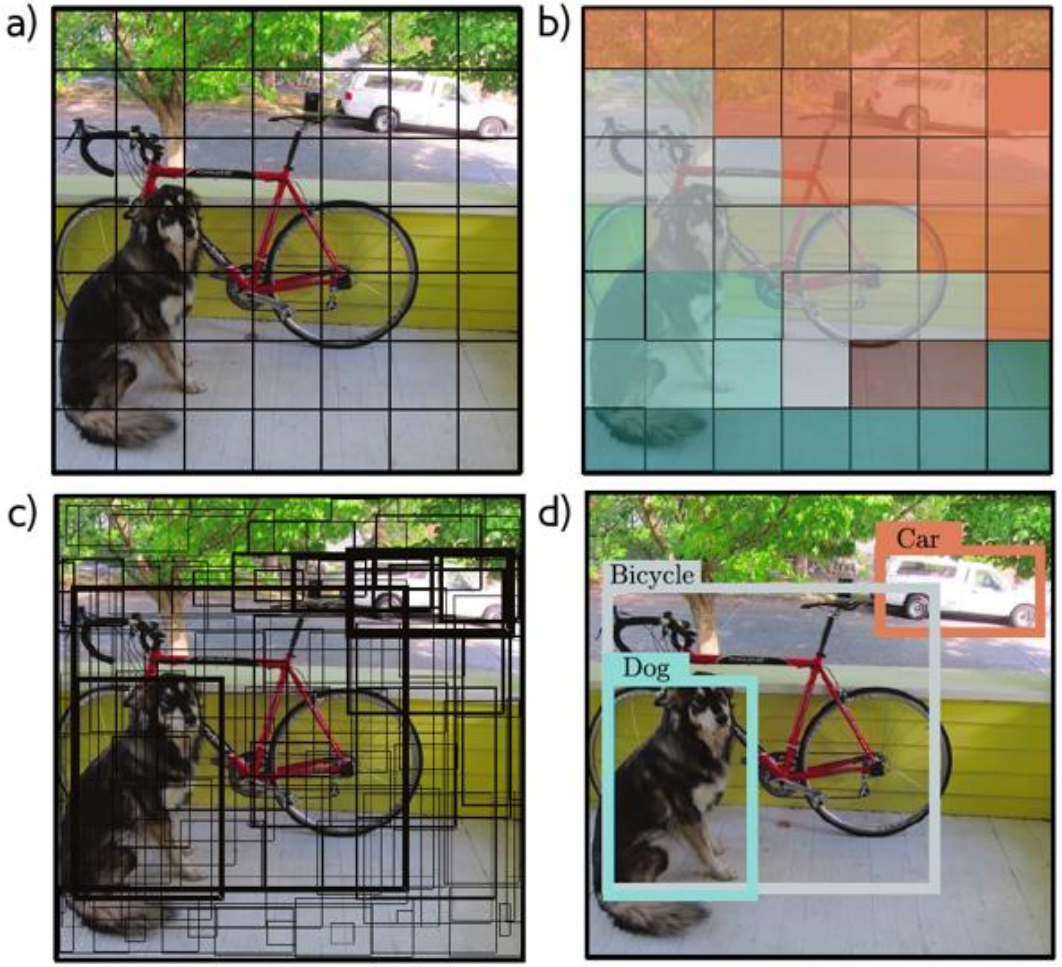
You Only Look Once (YOLO)



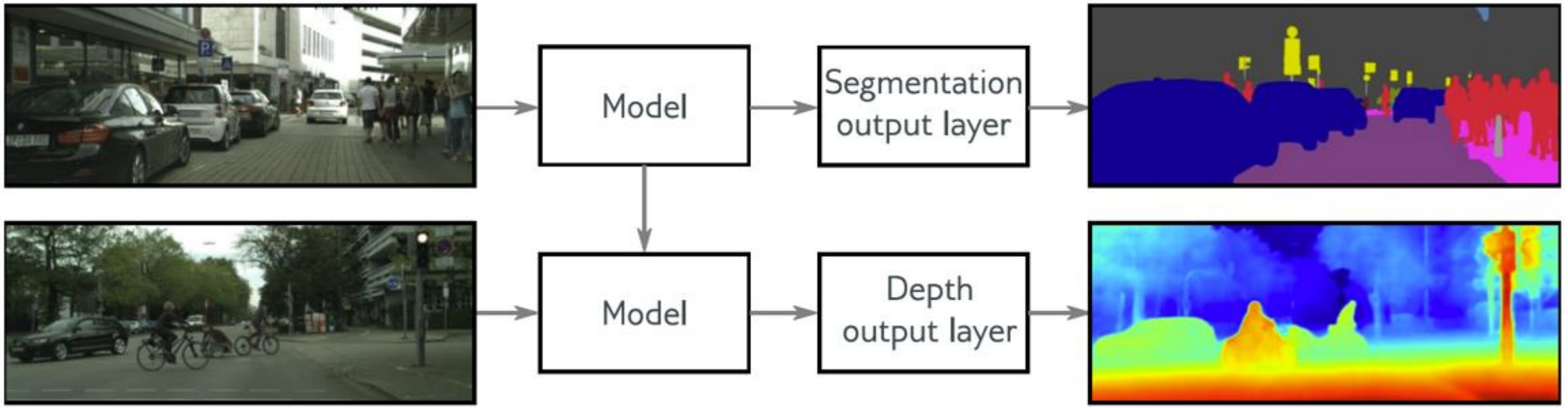
- Network similar to VGG (448x448 input)
- 7×7 grid of locations
- Predict class at each location
- Predict 2 bounding boxes at each location
  - Five parameters –x,y, height, width, and confidence
- Momentum, weight decay, dropout, and data augmentation
- Heuristic at the end to threshold and decide final boxes

2D CNNs for object detection

Object detection (YOLO)



2D CNNs for semantic segmentation



Transfer learning from ImageNet classification