CONVOLUTIONAL NEURAL NETWORKS: MOTIVATION & CONVOLUTION OPERATION

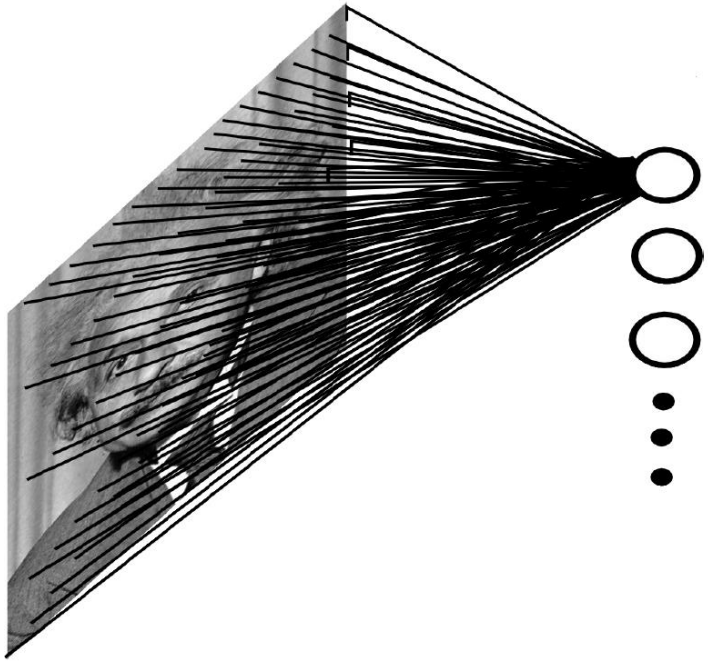
MOTIVATION

Fully connected neural network

- Example
  - 1000x1000 image
  - 1M hidden units

  ➔ $10^{12} (= 10^6 \times 10^6)$ parameters!

- Observation
  - Spatial correlation is local

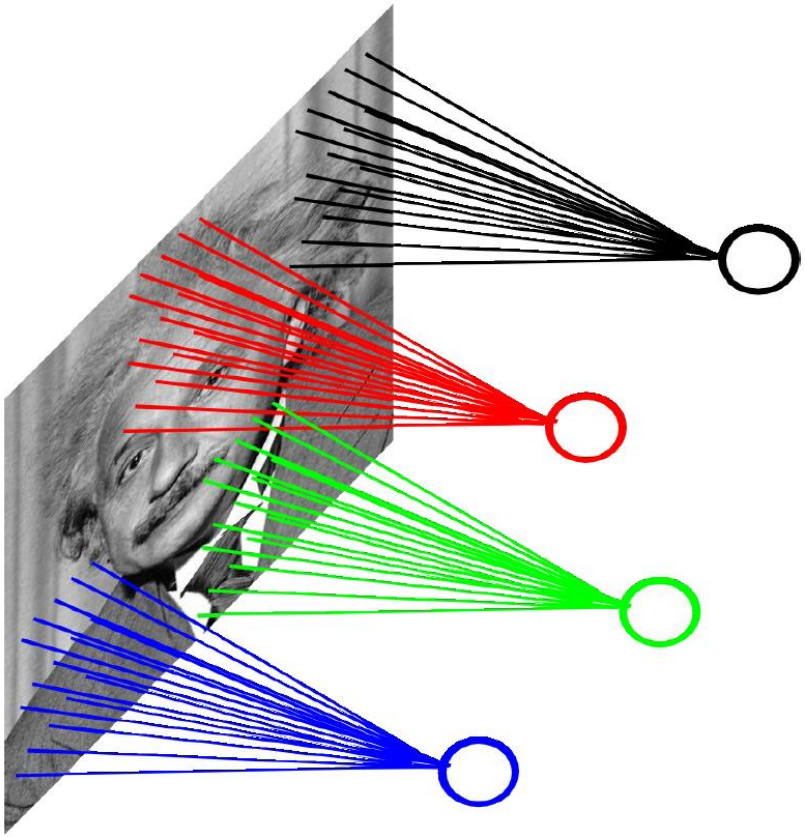


Locally connected neural net

- Example
  - 1000x1000 image
  - 1M hidden units
  - Filter size: 10x10

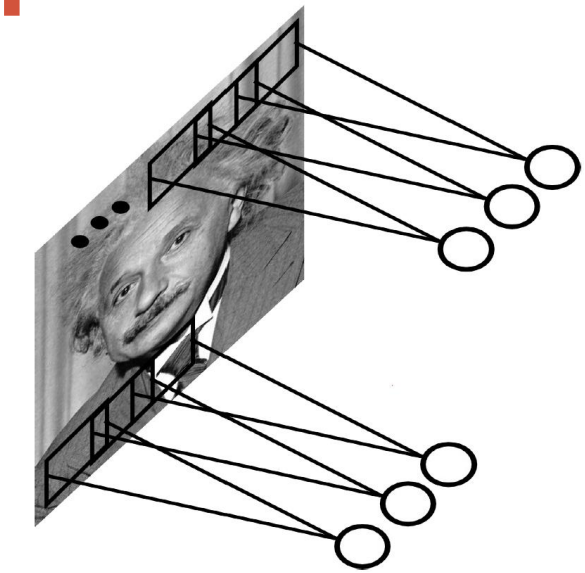
  ➔ $10^8 (= 10^6 \times 10 \times 10)$ parameters!

- Observation
  - Statistics is similar at different locations



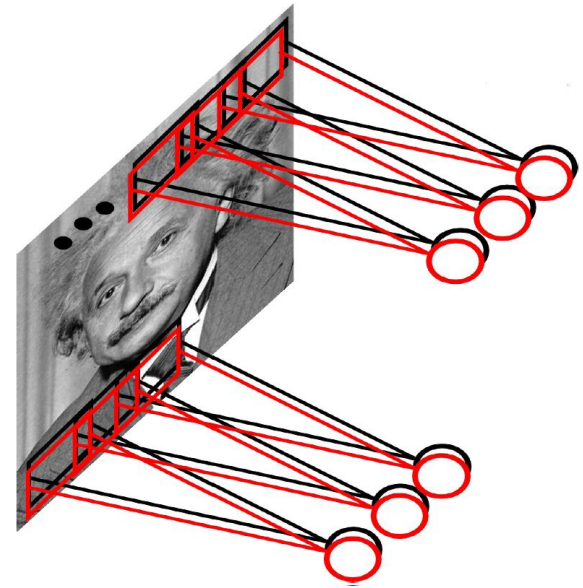
Convolution network

- Share the same parameters across different locations
  - Convolution with learned kernels



- Learn multiple filters
  - 1000x1000 image
  - 100 Filters
  - Filter size: 10x10

10,000 parameters



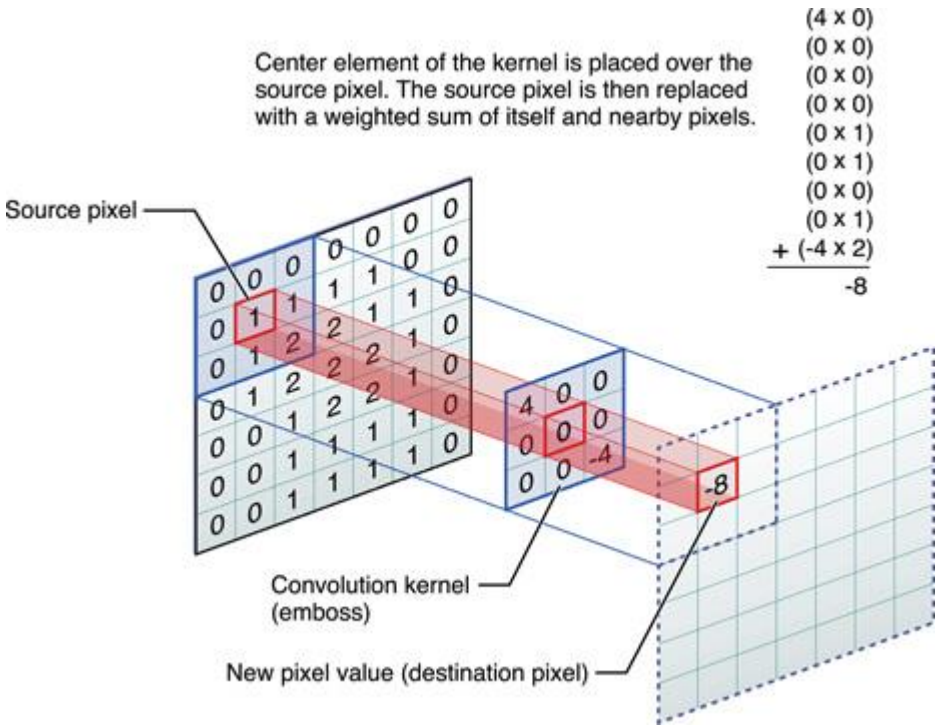
Convolution neural networks

- We can design neural networks that are specifically adapted for these problems
  - Must deal with very high-dimensional inputs
    - 1000x1000 pixels
  - Can exploit the 2D topology of pixels
  - Can build in invariance to certain variations we can expect
    - Translations, etc

- Ideas
  - Local connectivity
  - Parameter sharing

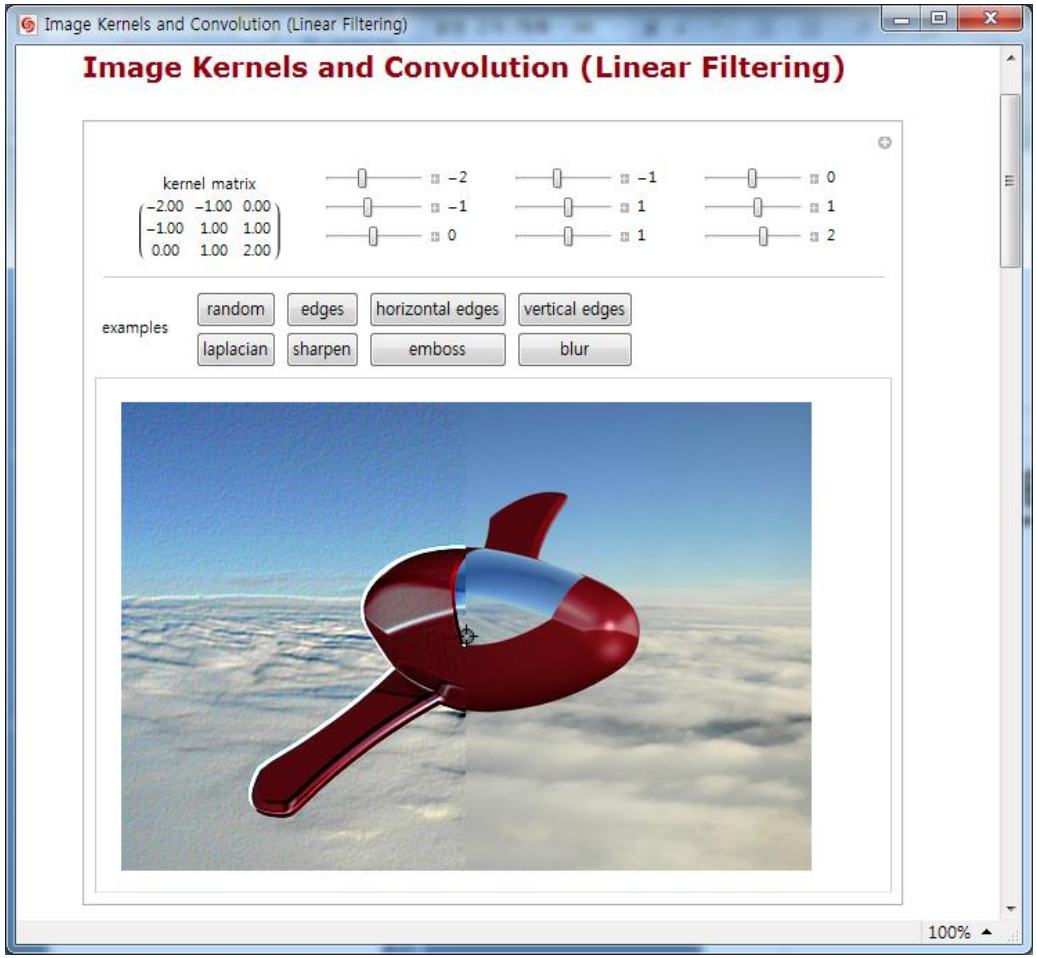
CONVOLUTION (IMAGE PROCESSING)

Convolution

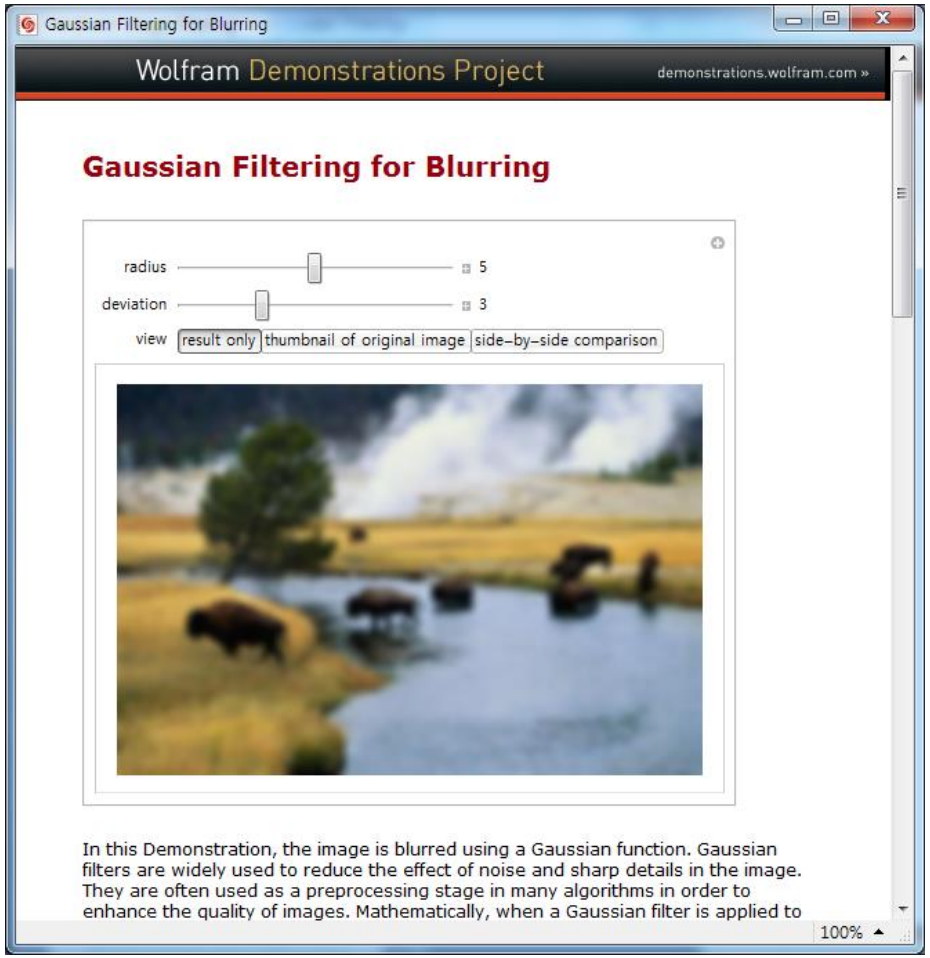


from: https://developer.apple.com/library/ios/documentation/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html

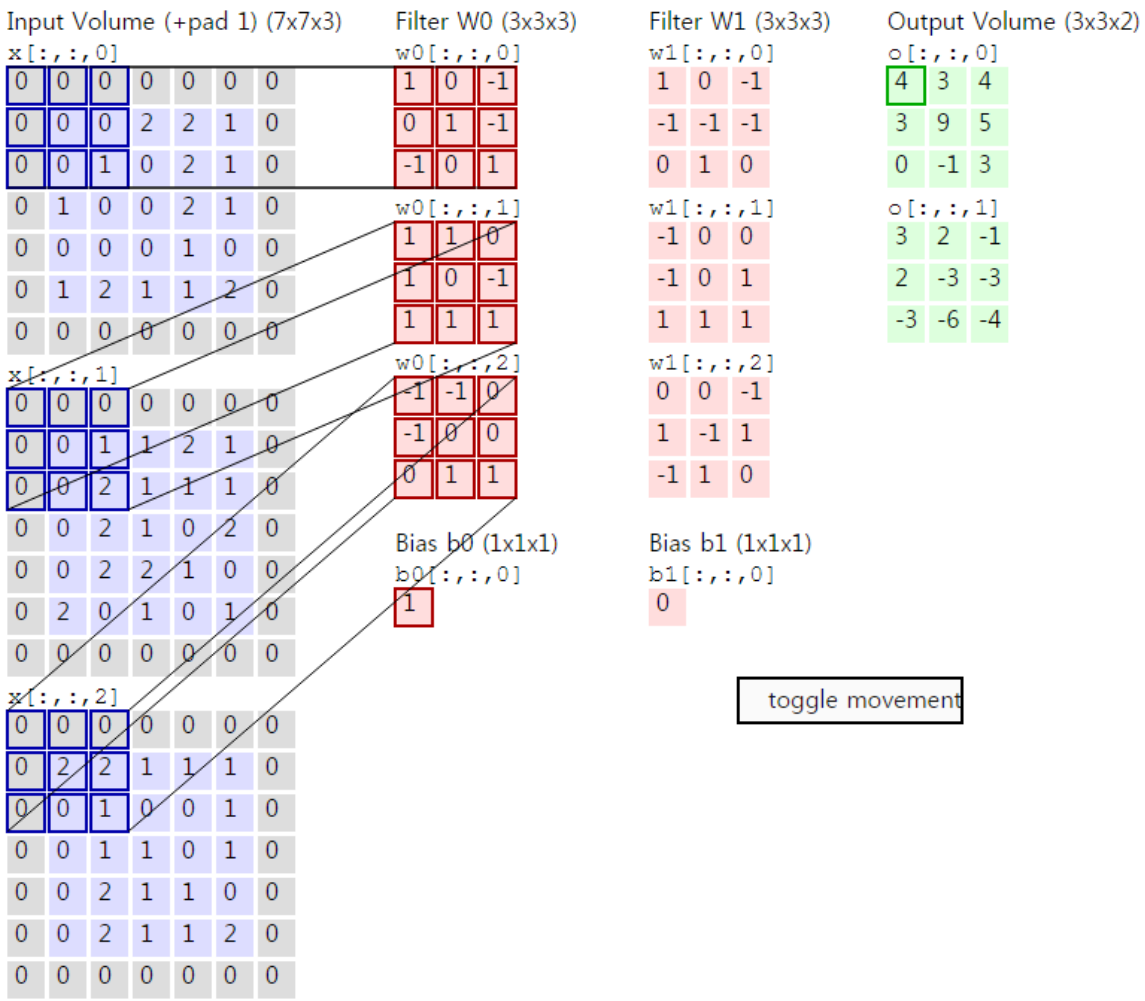
Linear filter

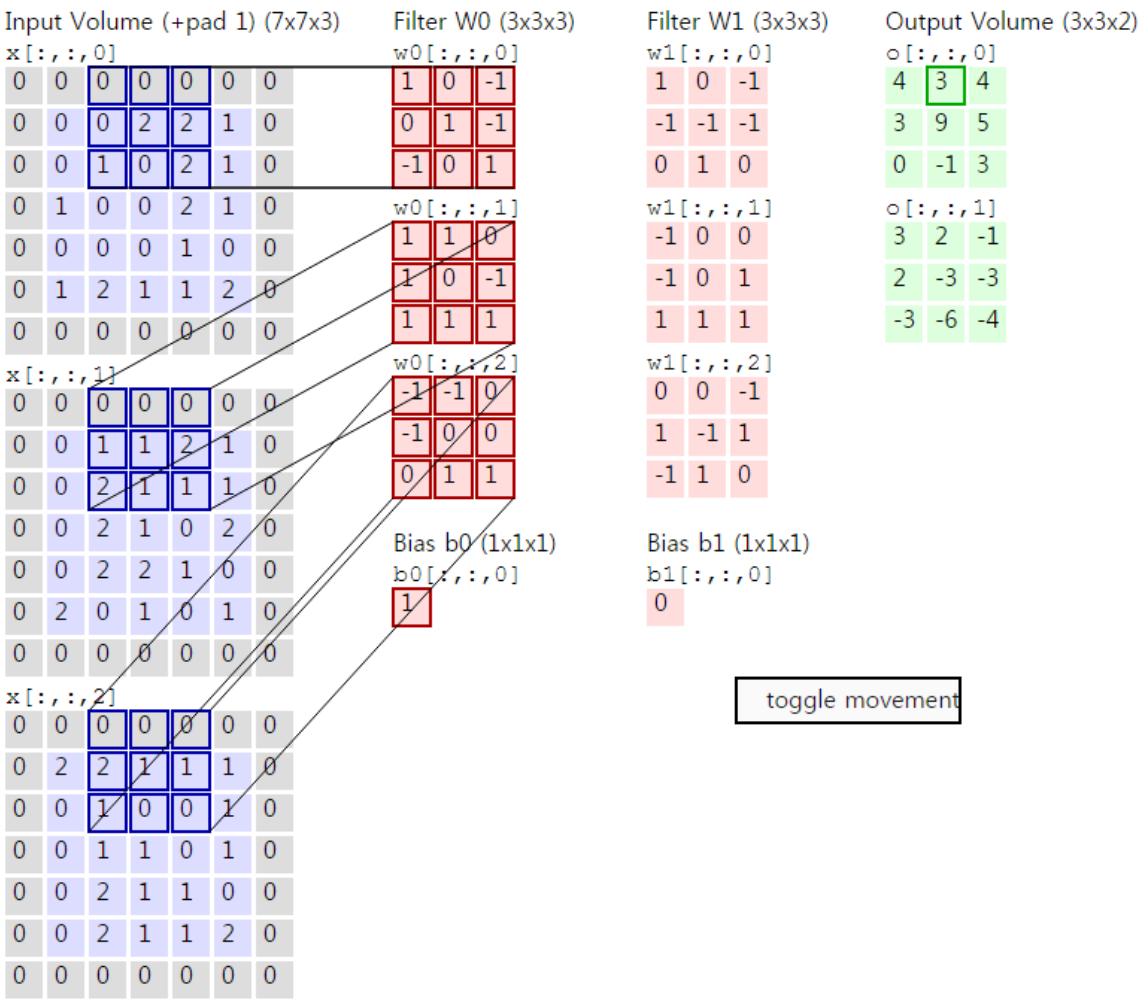


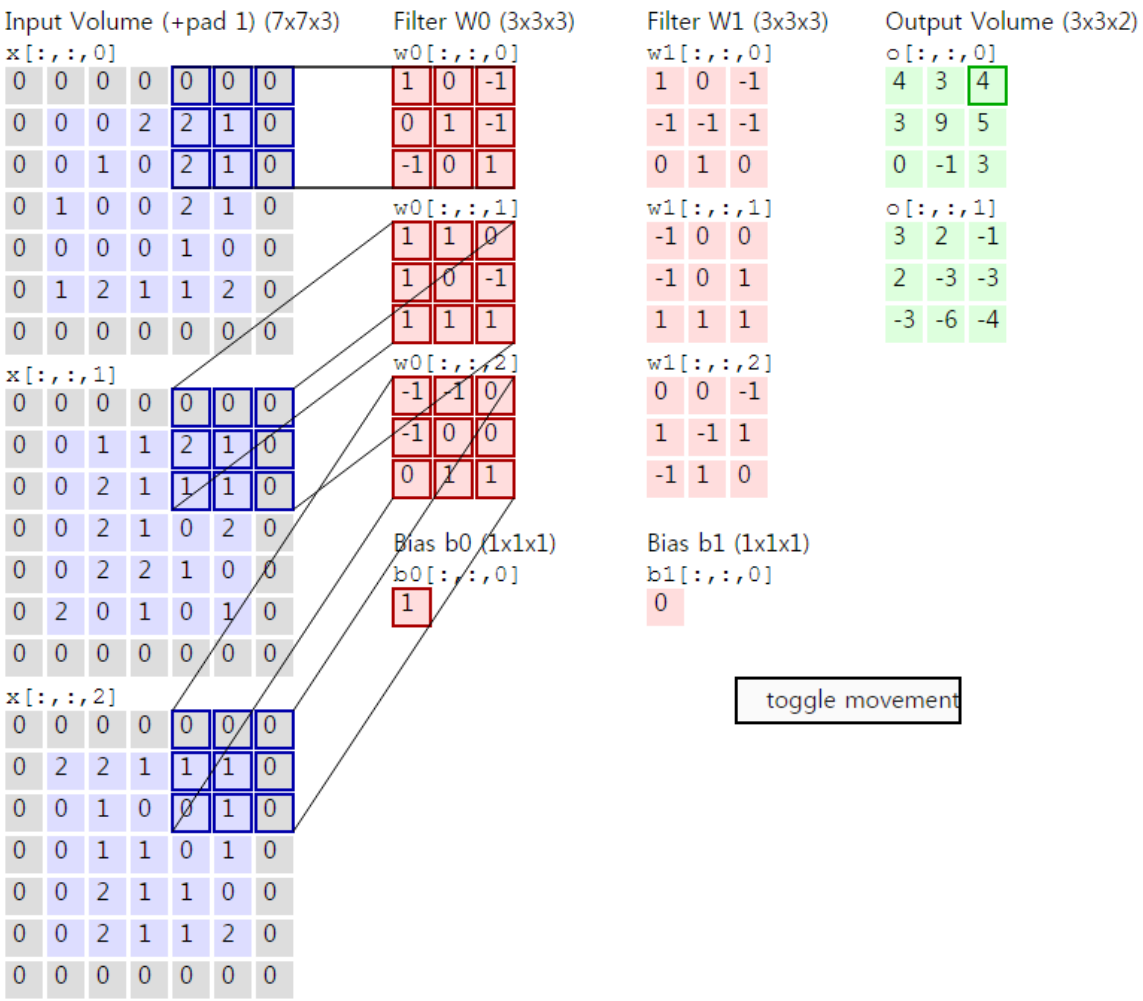
Linear filter (Gaussian)

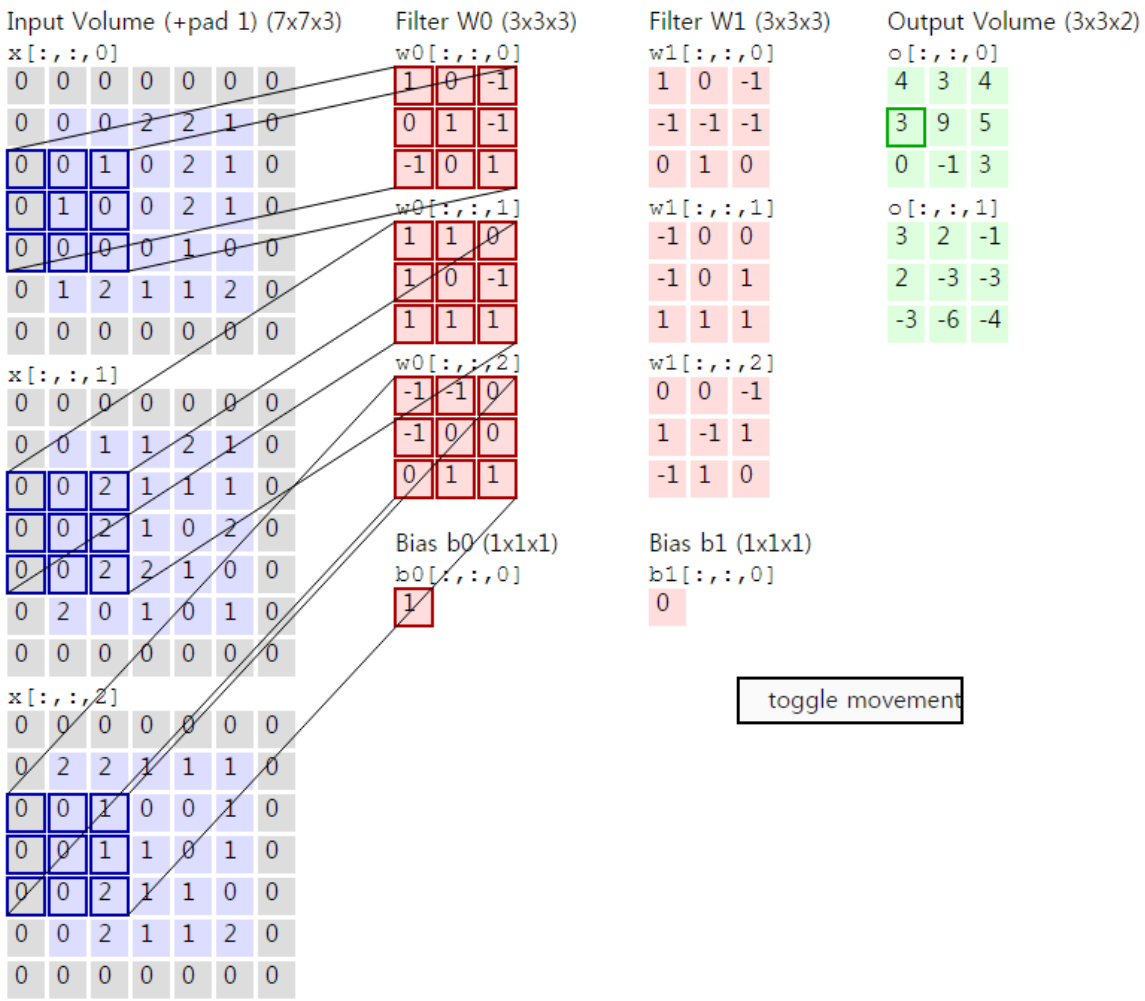


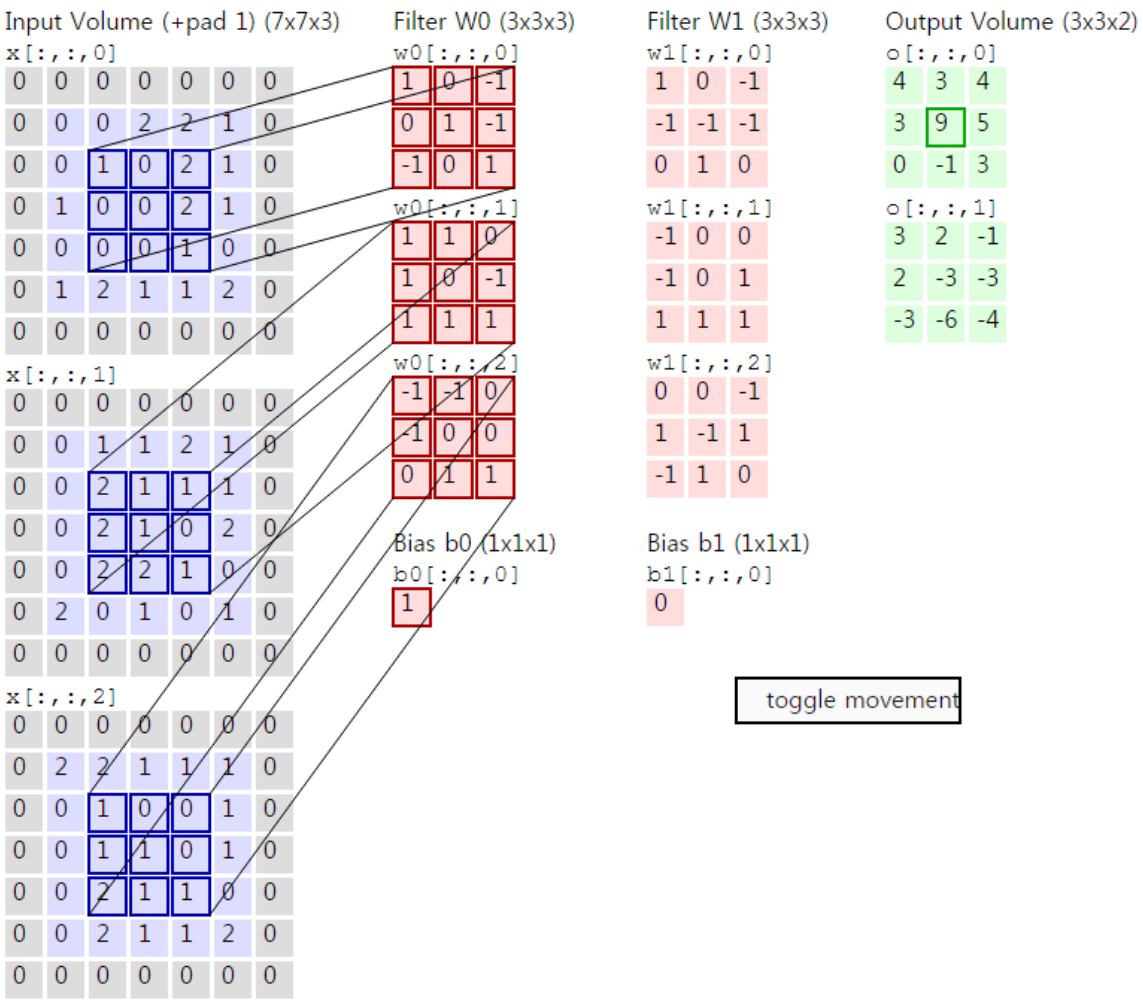
CONVOLUTION (DEEP LEARNING)

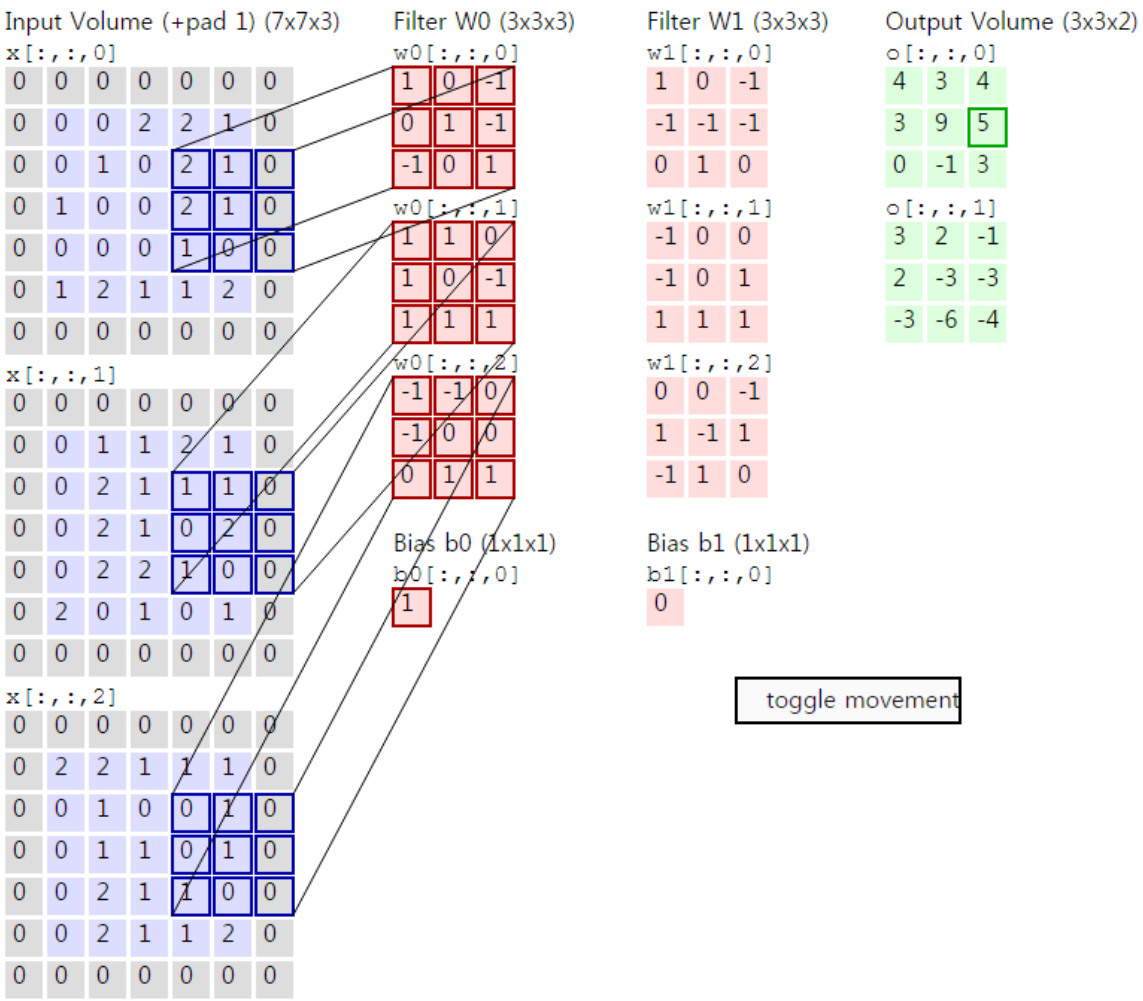

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

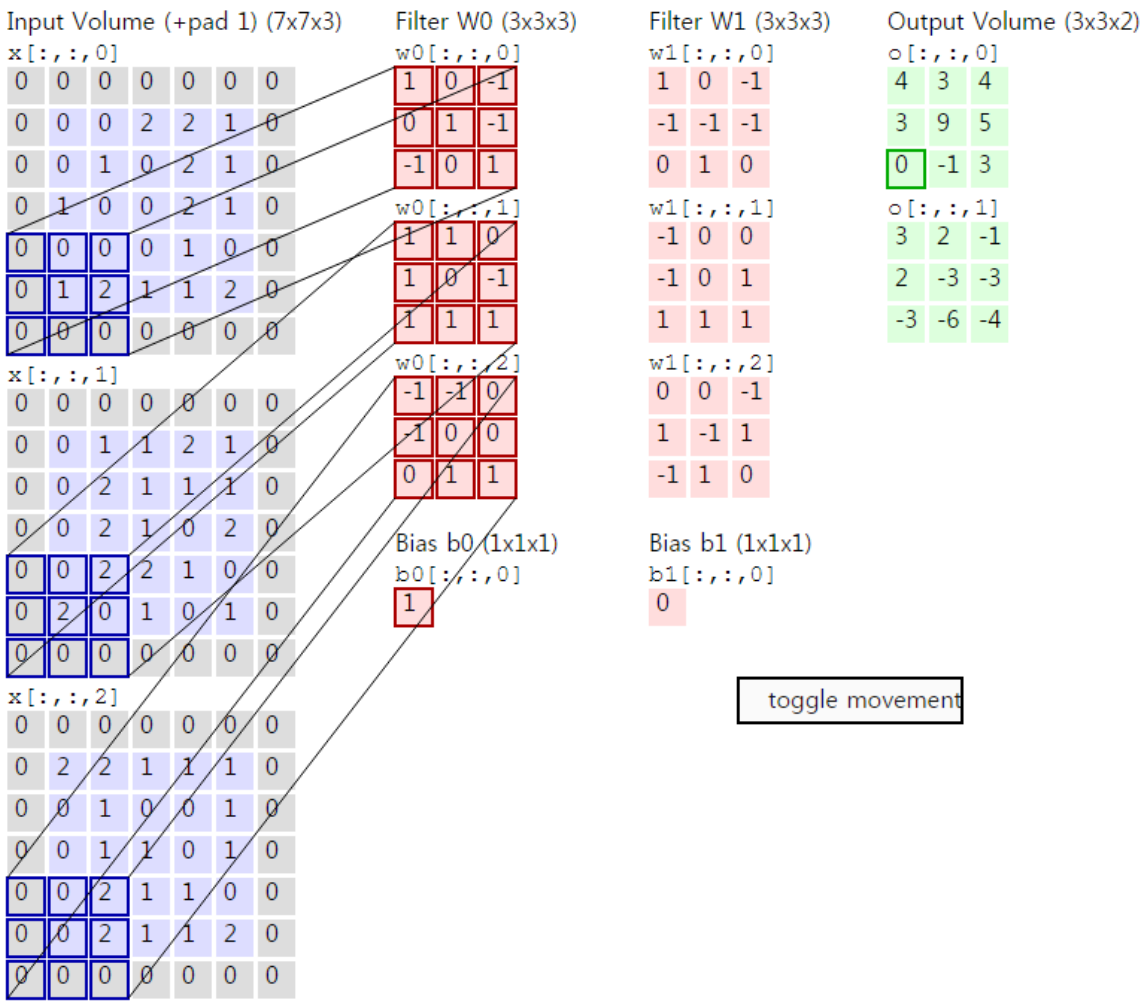

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

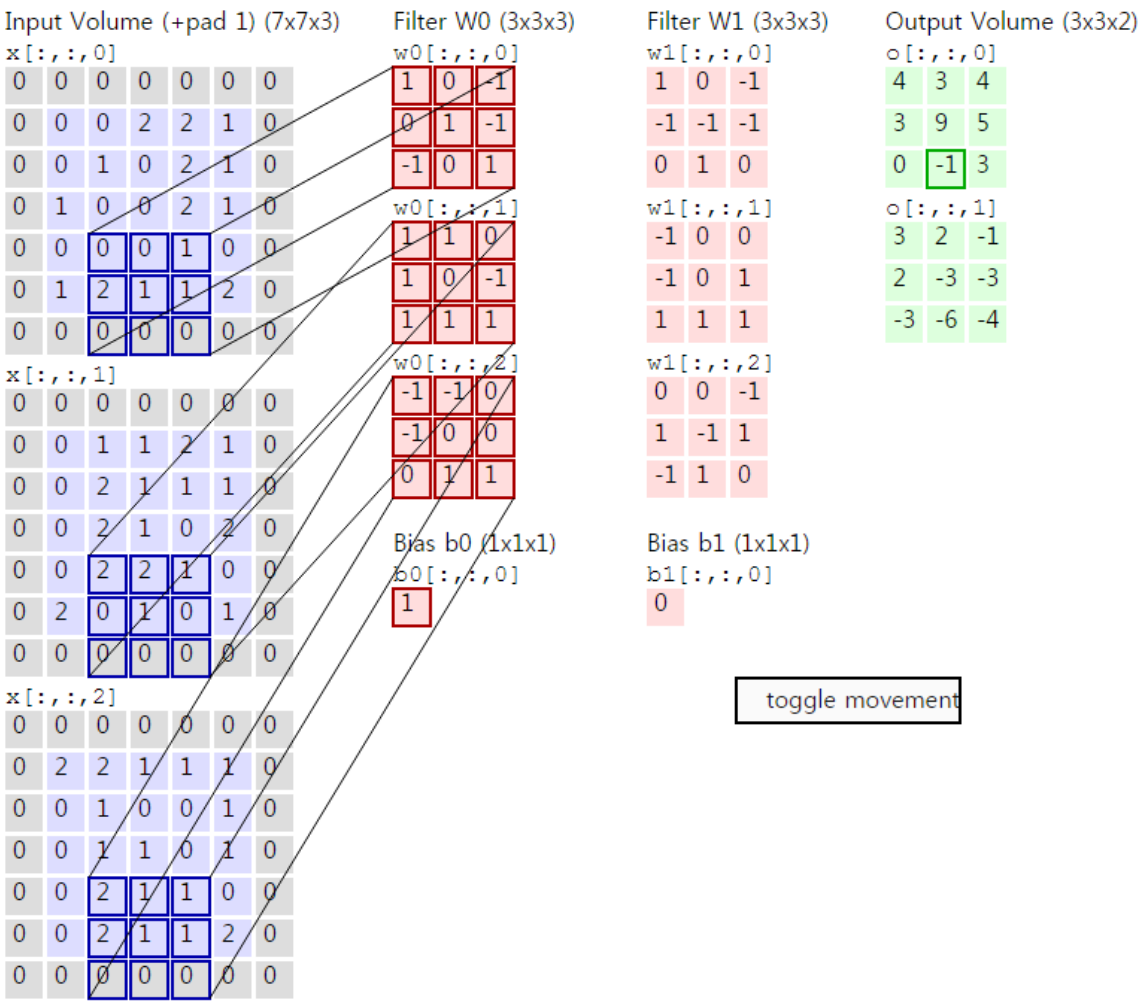

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

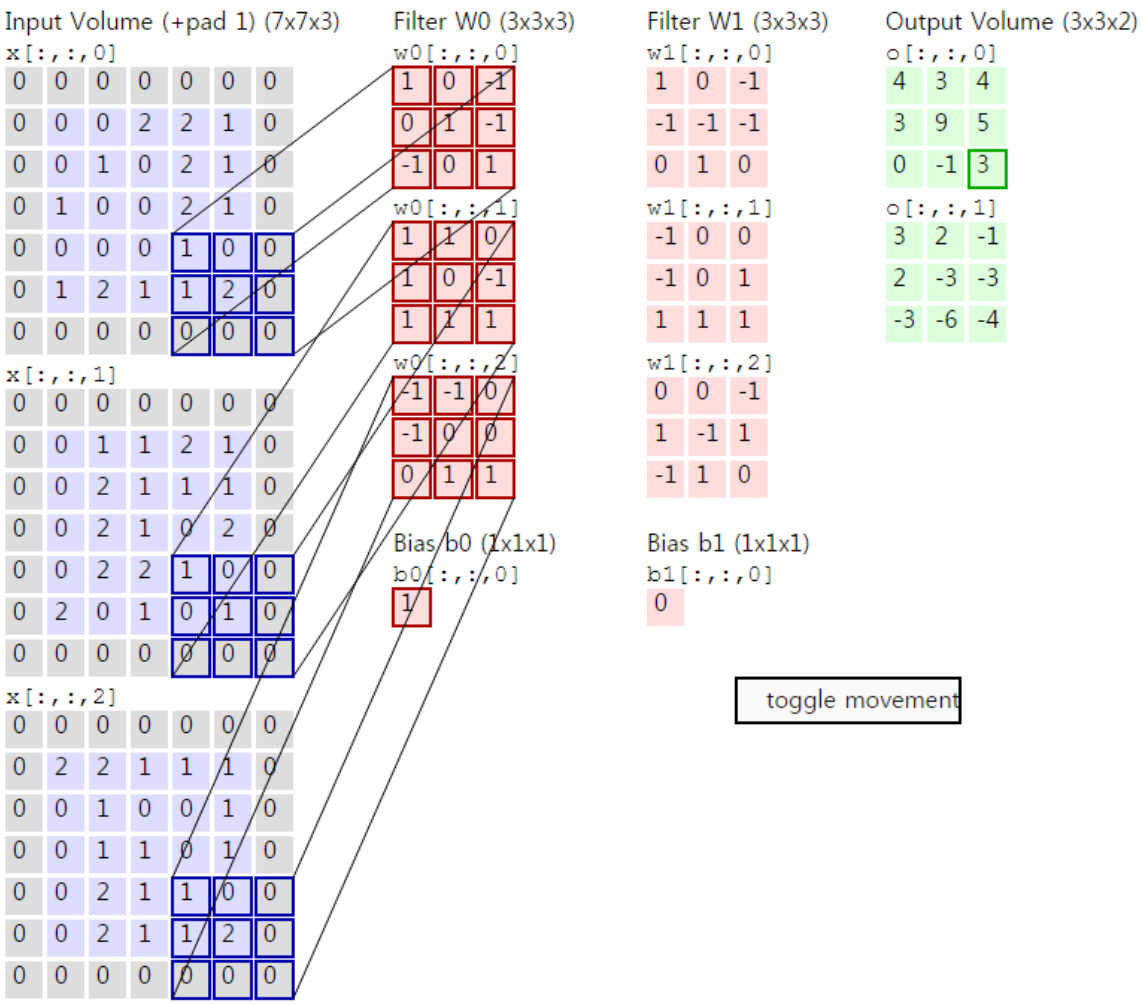

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

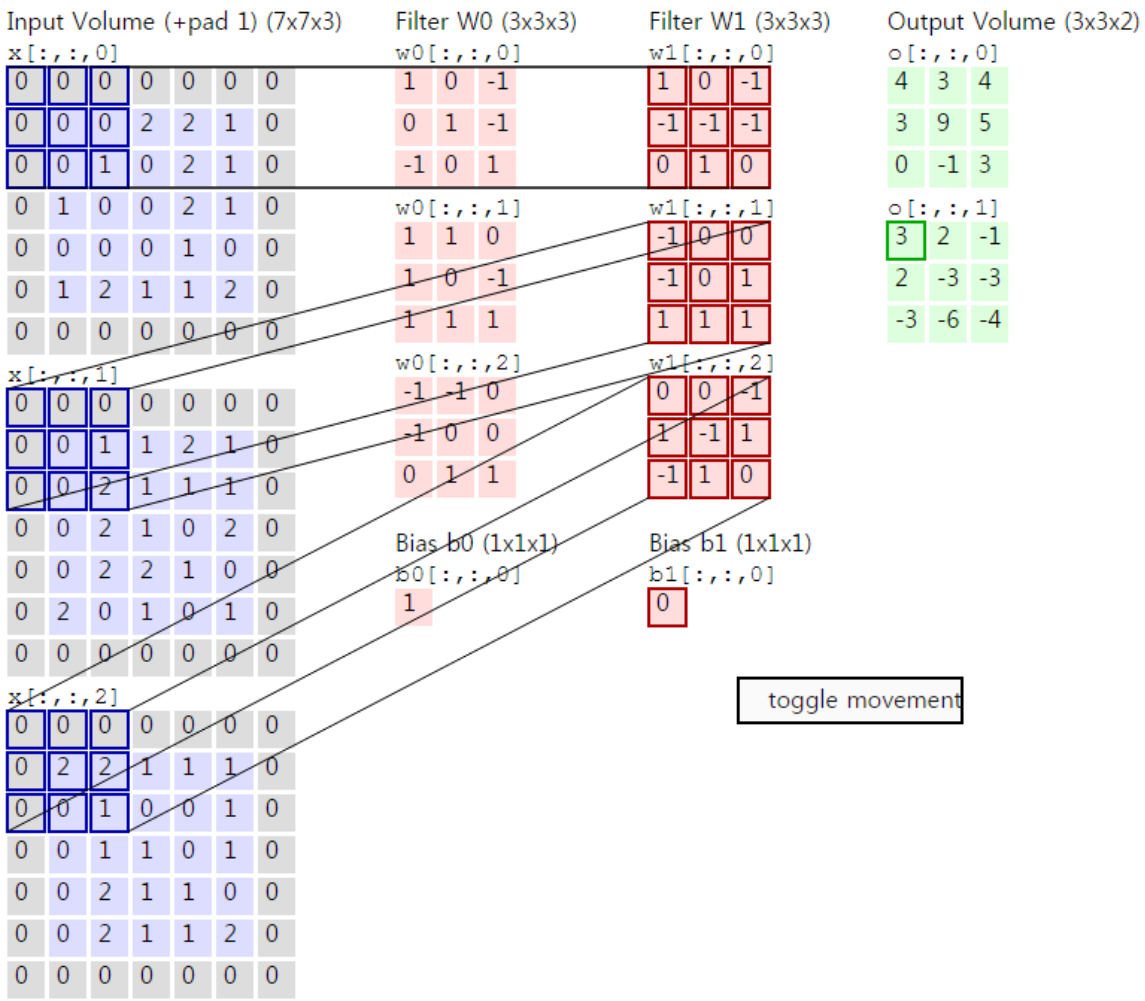

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

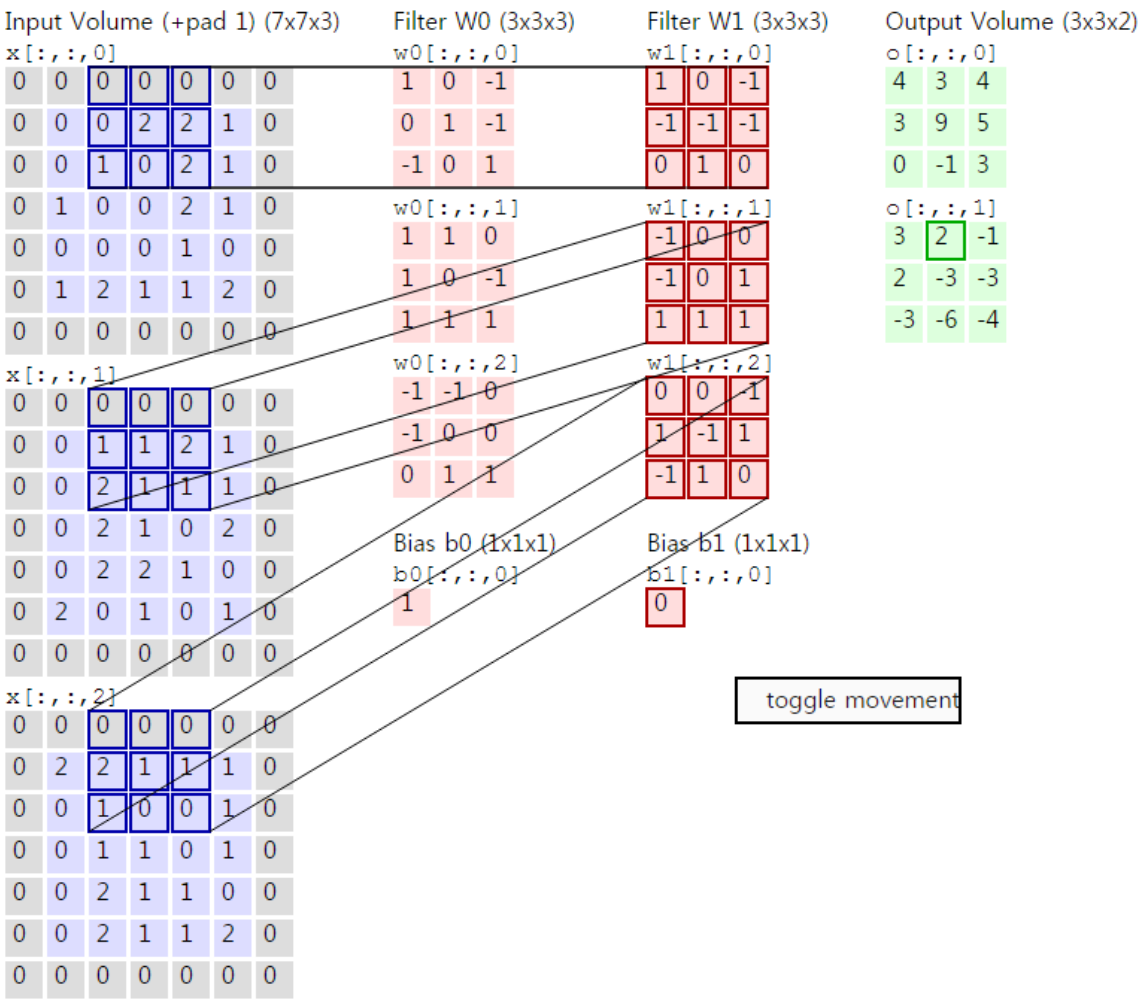

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

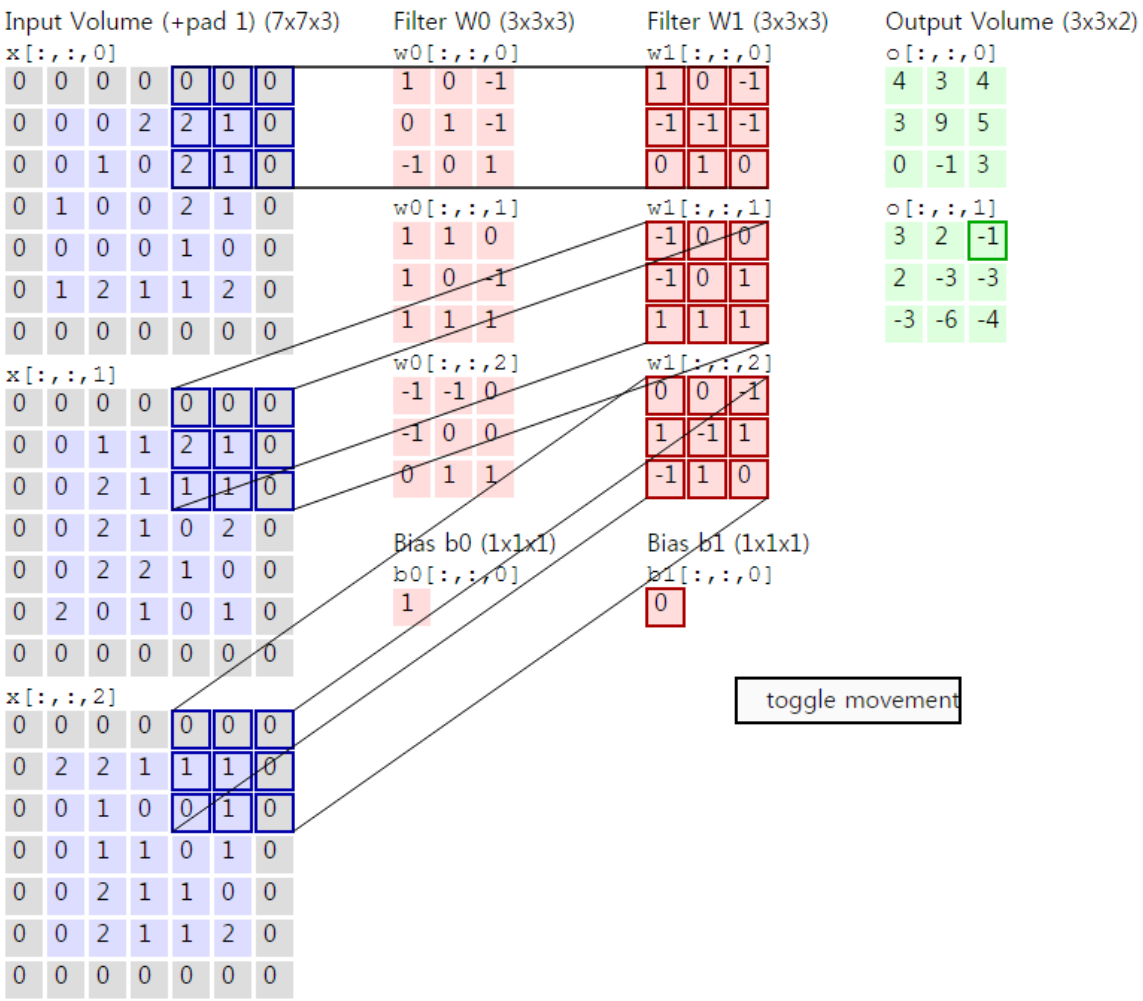

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

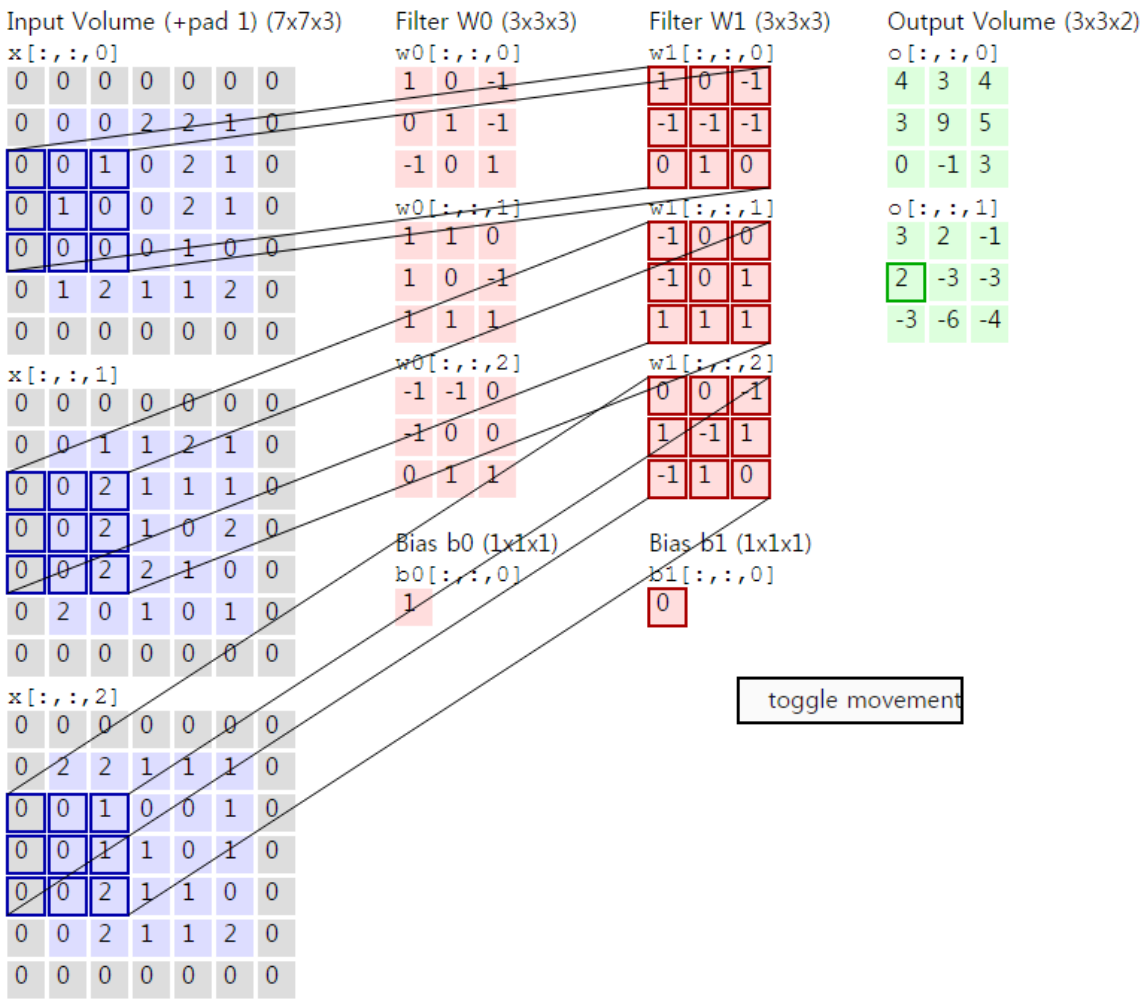

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

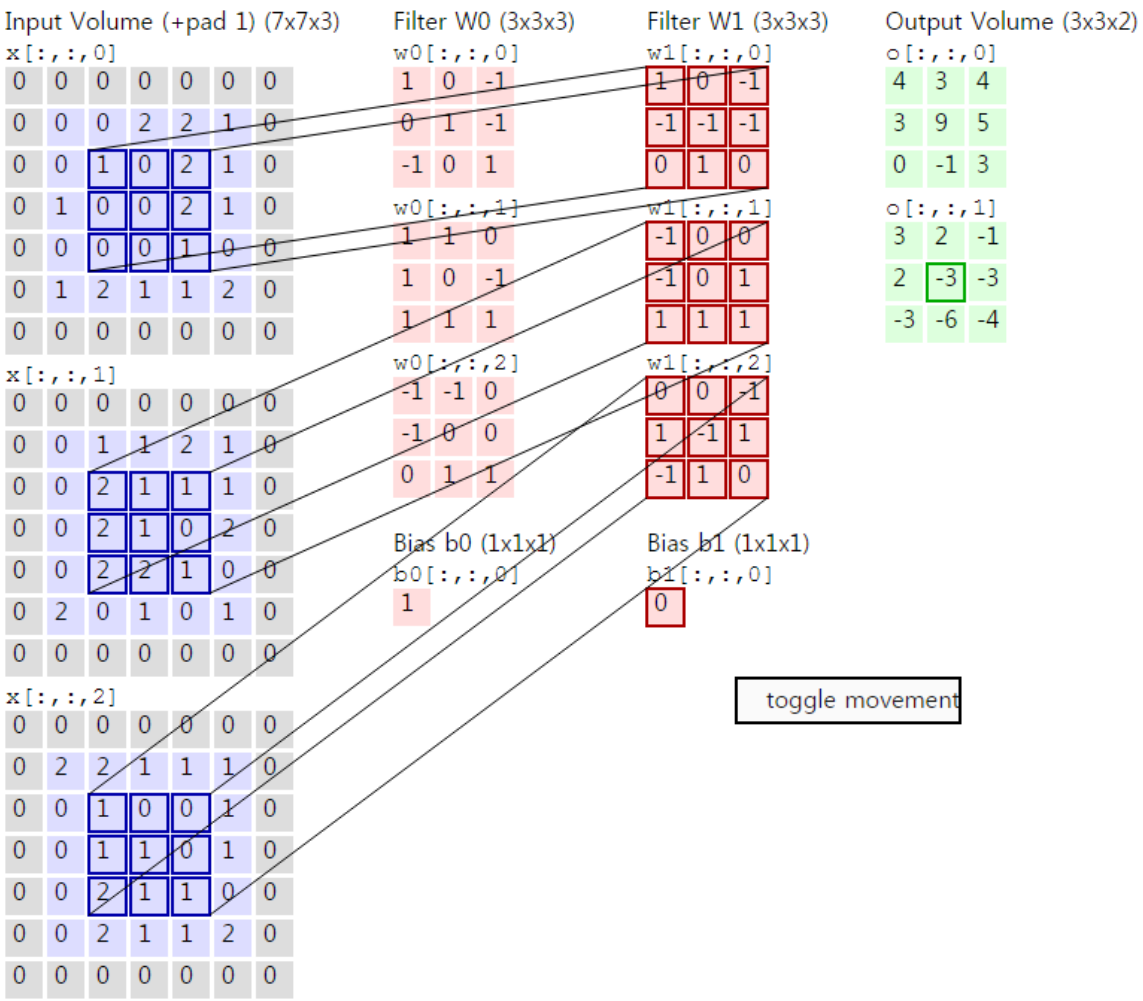

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

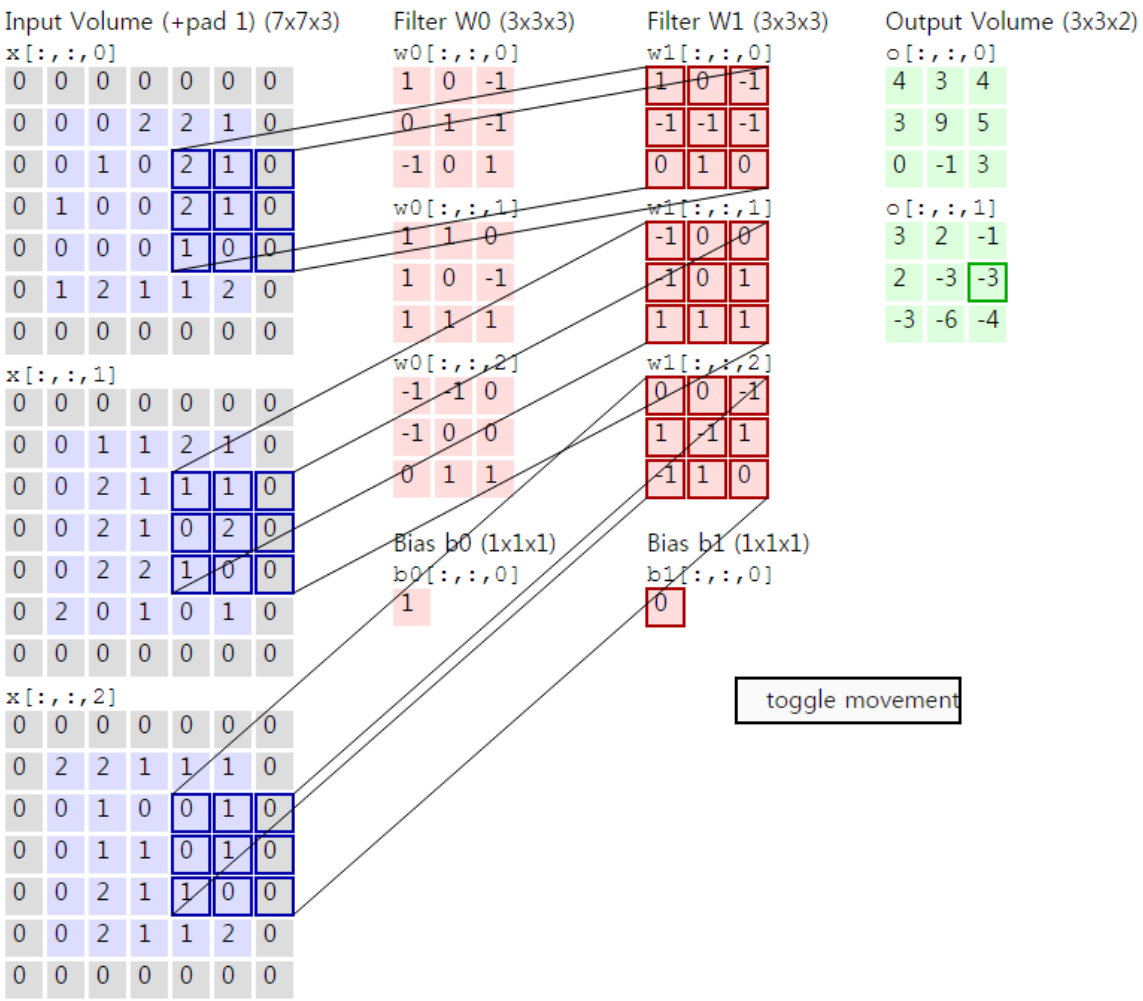

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

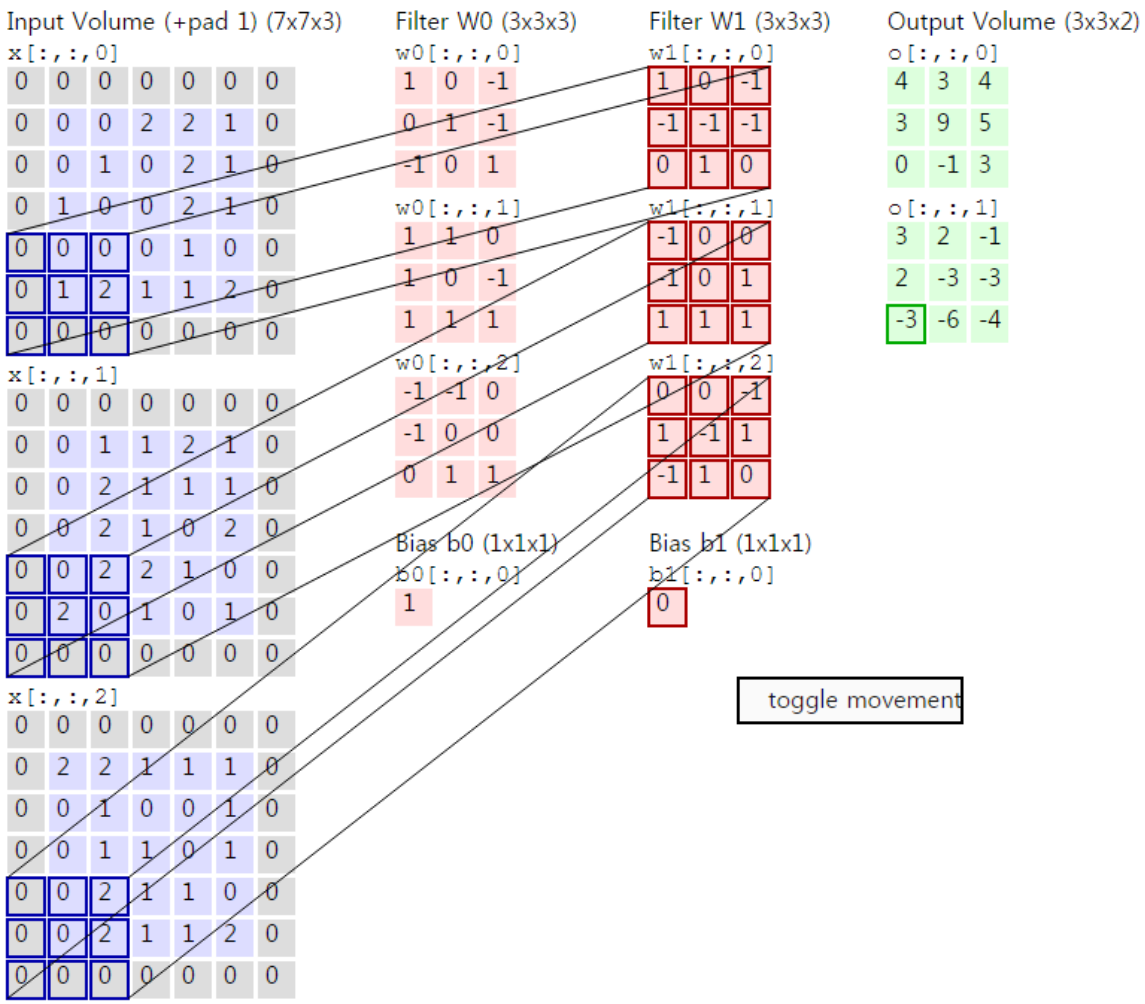

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

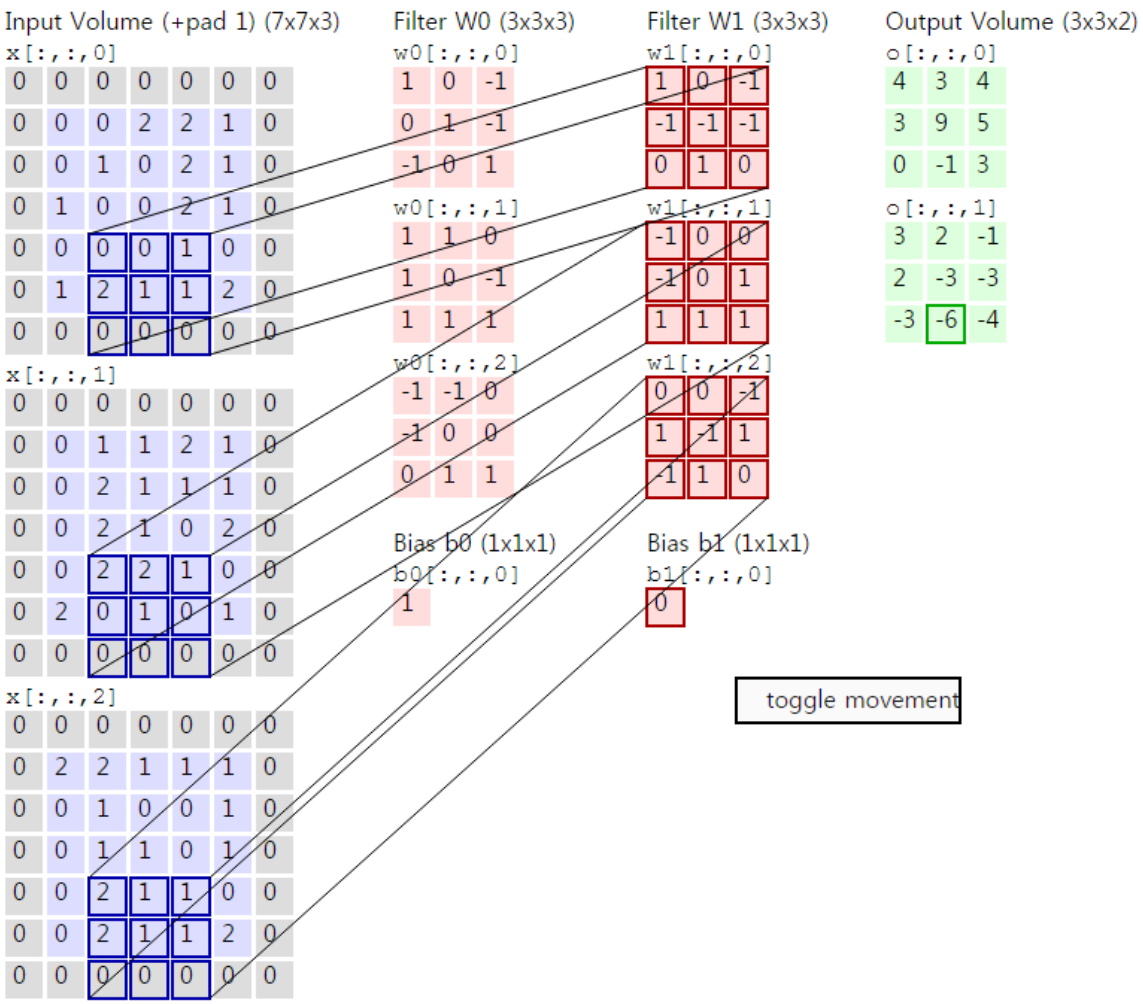

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

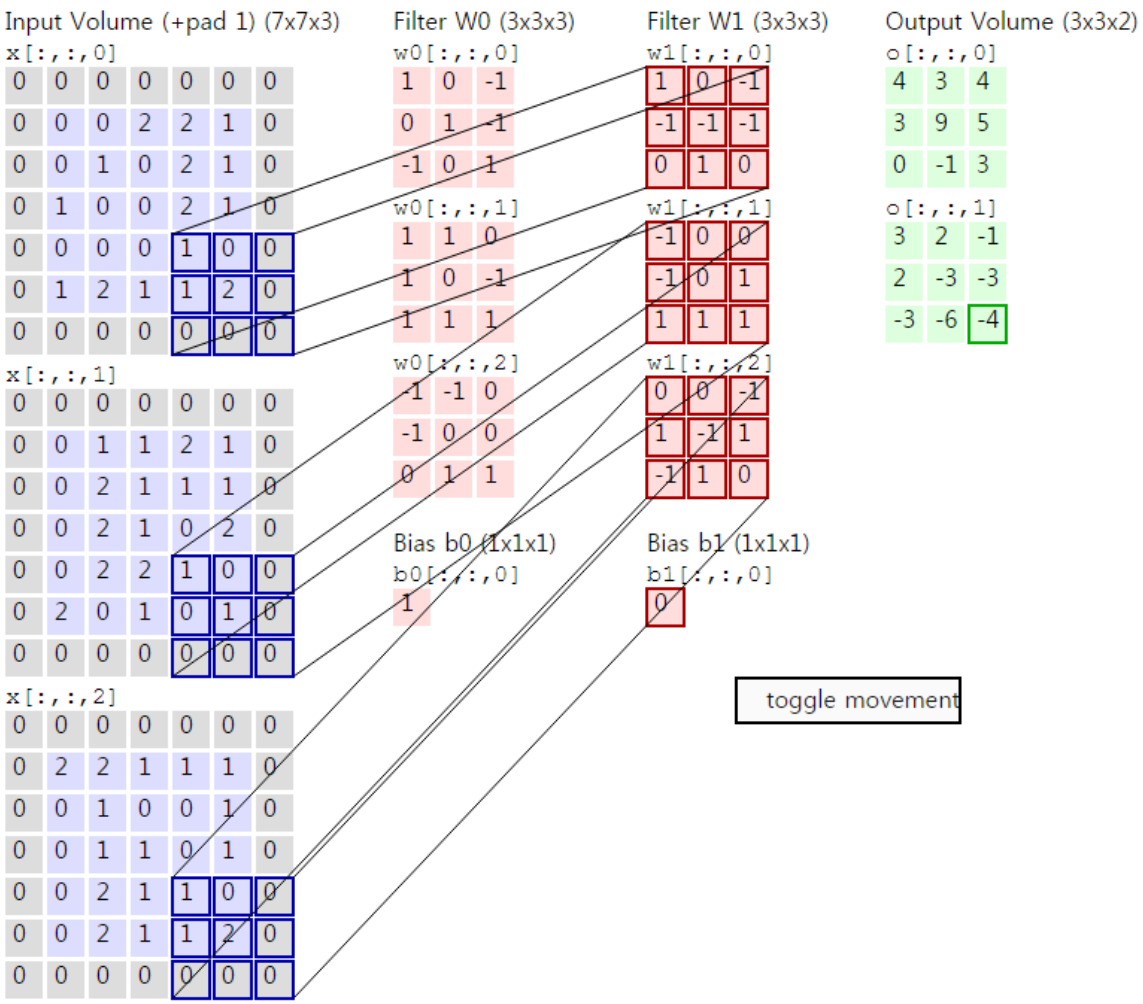

Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement


Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement


Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement


Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement


Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement


Input Volume (+pad 1) (7x7x3)
x[:,:,0]
0 0 0 0 0 0 0
0 0 0 2 2 1 0
0 0 1 0 2 1 0
0 1 0 0 2 1 0
0 0 0 0 1 0 0
0 1 2 1 1 2 0
0 0 0 0 0 0 0
x[:,:,1]
0 0 0 0 0 0 0
0 0 1 1 2 1 0
0 0 2 1 1 1 0
0 0 2 1 0 2 0
0 0 2 2 1 0 0
0 2 0 1 0 1 0
0 0 0 0 0 0 0
x[:,:,2]
0 0 0 0 0 0 0
0 2 2 1 1 1 0
0 0 1 0 0 1 0
0 0 1 1 0 1 0
0 0 2 1 1 0 0
0 0 2 1 1 2 0
0 0 0 0 0 0 0
Filter W0 (3x3x3)
w0[:,:,0]
1 0 -1
0 1 -1
-1 0 1
w0[:,:,1]
1 1 0
1 0 -1
1 1 1
w0[:,:,2]
-1 -1 0
-1 0 0
0 1 1
Bias b0 (1x1x1)
b0[:,:,0]
1
Filter W1 (3x3x3)
w1[:,:,0]
1 0 -1
-1 -1 -1
0 1 0
w1[:,:,1]
-1 0 0
-1 0 1
1 1 1
w1[:,:,2]
0 0 -1
1 -1 1
-1 1 0
Bias b1 (1x1x1)
b1[:,:,0]
0
Output Volume (3x3x2)
o[:,:,0]
4 3 4
3 9 5
0 -1 3
o[:,:,1]
3 2 -1
2 -3 -3
-3 -6 -4
toggle movement

ALEXNET

THE IMAGENET LARGE SCALE VISUAL RECOGNITION CHALLENGE (ILSVRC)

Backpack

Flute

Strawberry

Traffic light

Backpack

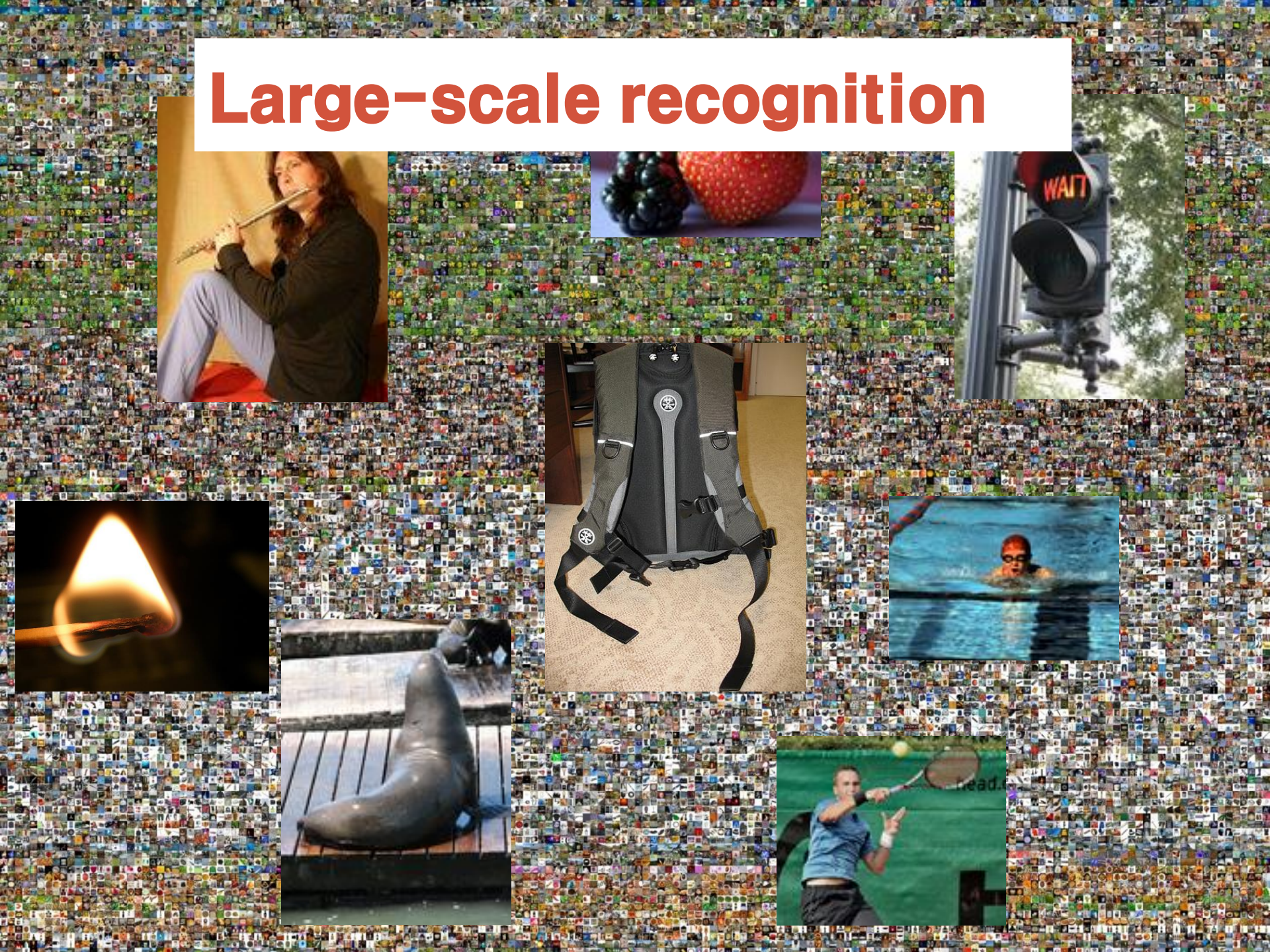
Matchstick

Sea lion

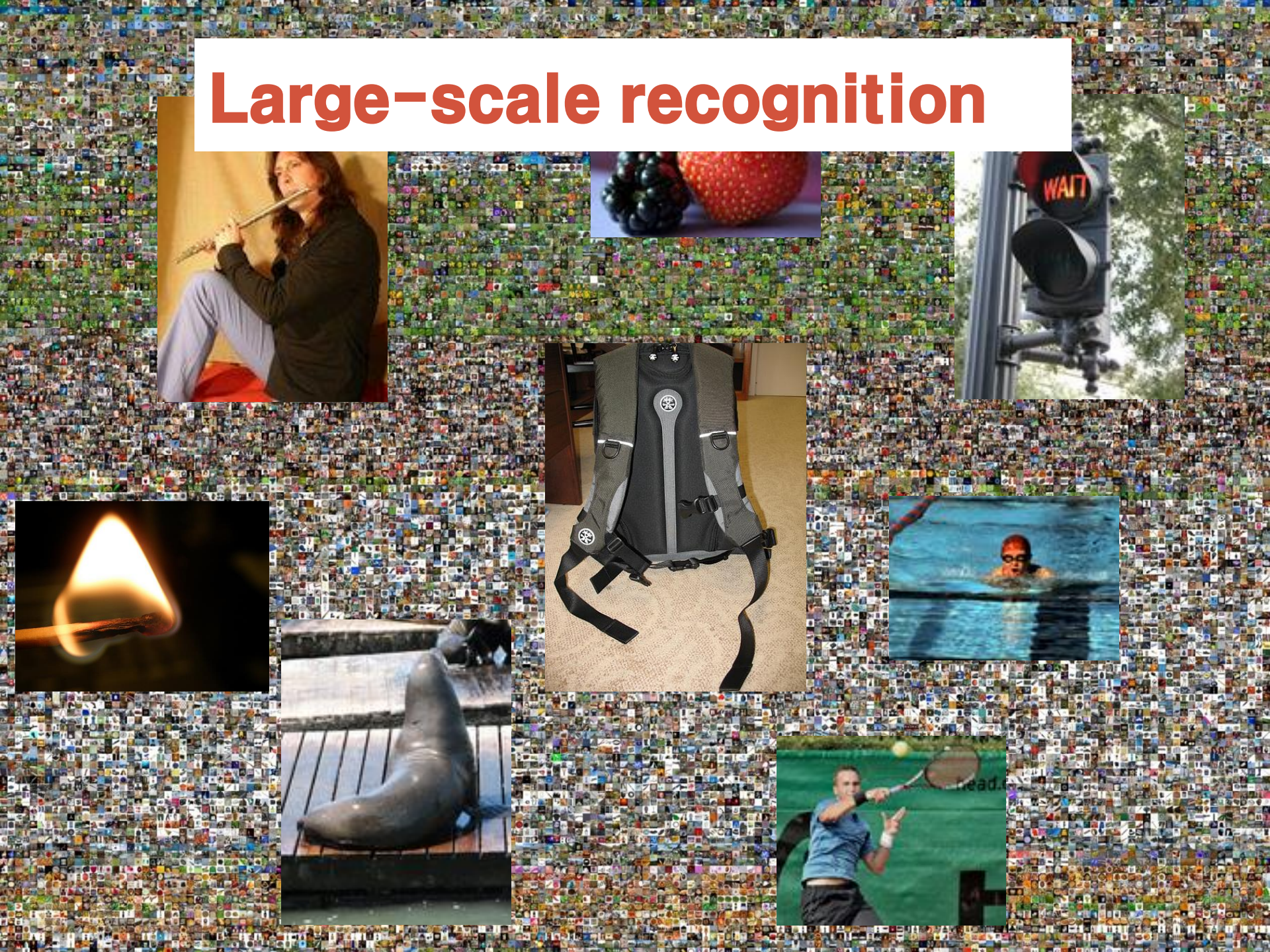
Bathing cap

Racket

Large-scale recognition

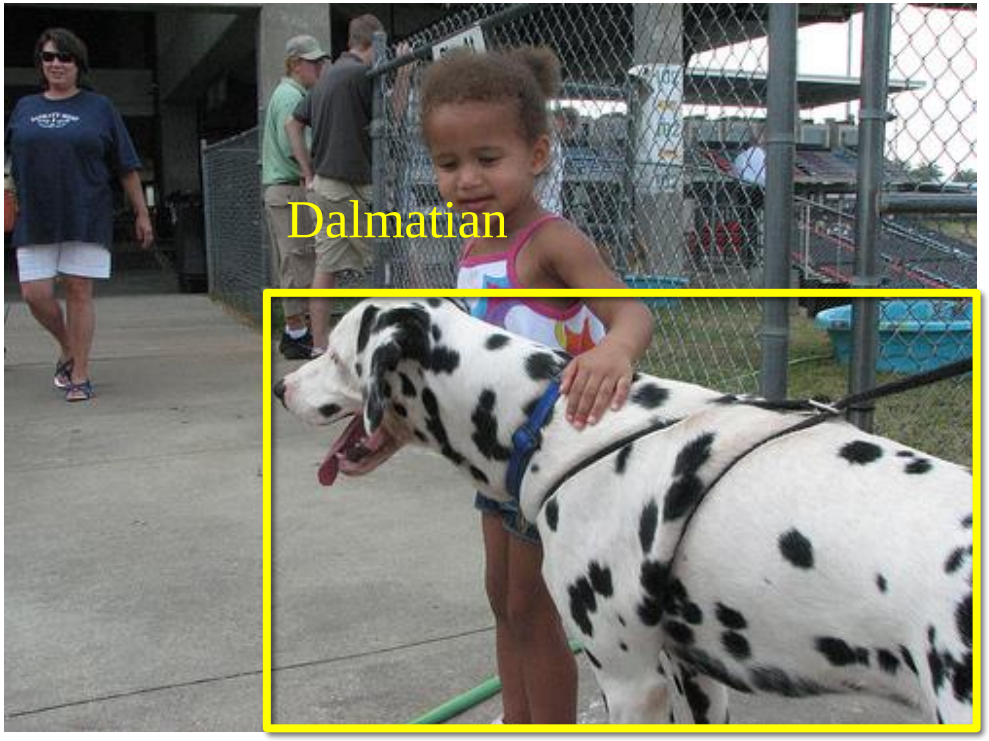


Large-scale recognition



Large Scale Visual Recognition Challenge (ILSVRC) 2010-2012

1000 object classes **1,431,167 images**



http://image-net.org/challenges/LSVRC/{2010,2011,2012}

Variety of object classes in ILSVRC

ILSVRC Task 1: Classification

Steel drum

ILSVRC Task 1: Classification

Steel drum

Output:
Scale
T-shirt
<u>Steel drum</u>
Drumstick
Mud turtle

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

ILSVRC Task 1: Classification

Steel drum

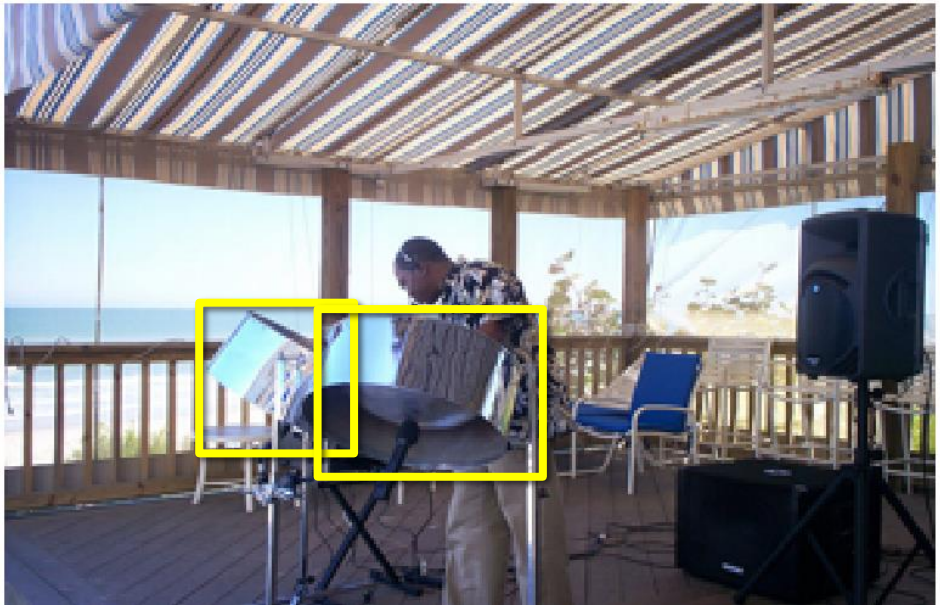
Output:
Scale
T-shirt
<u>Steel drum</u>
Drumstick
Mud turtle

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

$$\text{Accuracy} = \frac{1}{N} \sum_{\substack{N \\ \text{images}}} 1[\text{correct on image i}]$$

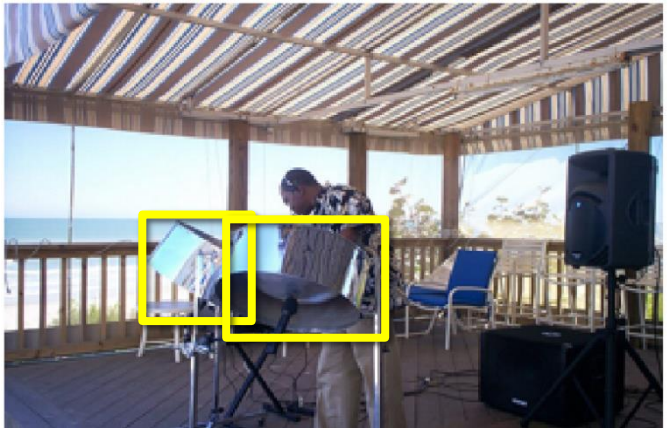
ILSVRC Task 2: Classification + Localization

Steel drum



ILSVRC Task 2: Classification + Localization

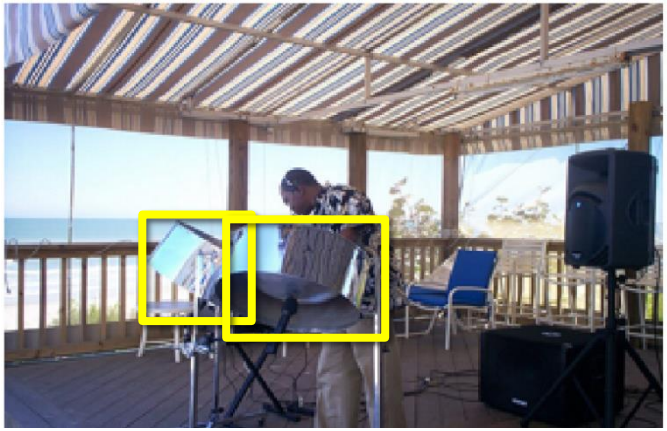
Steel drum



Output

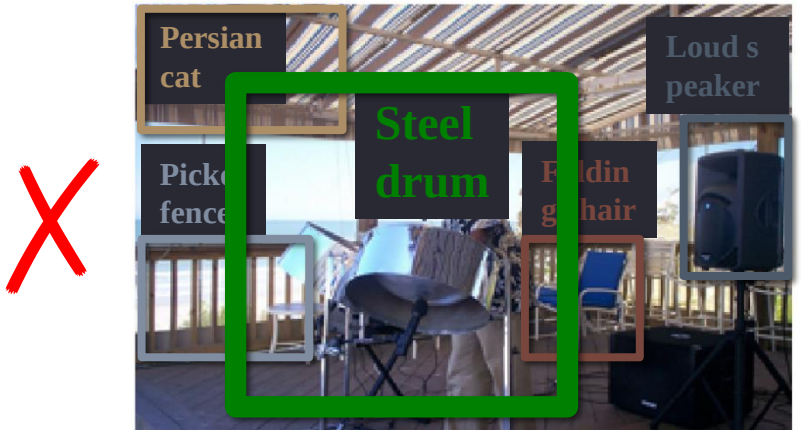
ILSVRC Task 2: Classification + Localization

Steel drum

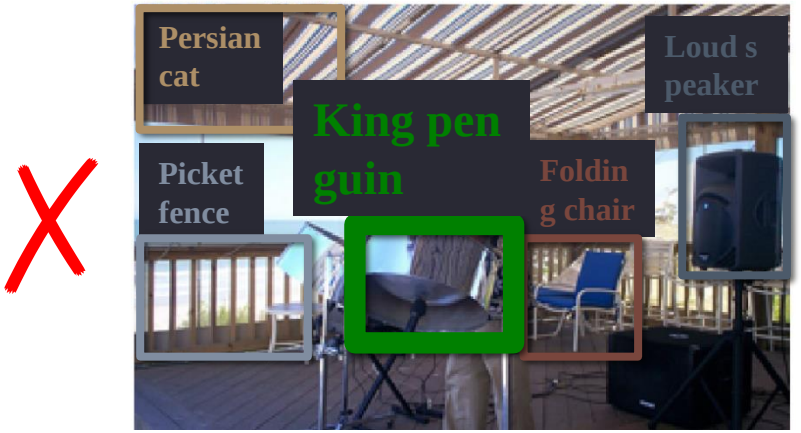


Output

Output (bad localization)

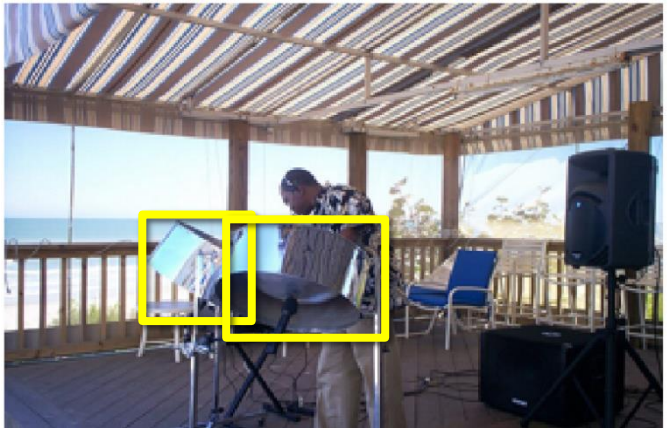


Output (bad classification)



ILSVRC Task 2: Classification + Localization

Steel drum



Output

$$\text{Accuracy} = \frac{1}{N} \sum_{\text{N-images}} 1[\text{correct on image i}]$$

Classification: Comparison

Submission	Method	Error rate
SuperVision	Deep CNN	0.16422
ISI	FV: SIFT, LBP, GIST, CSIFT	0.26172
XRCE/INRIA	FV: SIFT and color 1M-dim features	0.27058
OXFORD_VGG	FV: SIFT and color 270K-dim features	0.27302

Classification + Localization

Team name	Filename	Error (5 guesses)	Description
SuperVision	test-rect-preds-144-cloc-141-146.2009-131-137-145-	0.335463	Using extra training data for classification from ImageNet Fall 2011 release
SuperVision	test-rect-preds-144-cloc-131-137-145-135-145f.txt	0.341905	Using only supplied training data
OXFORD_VGG	test_adhocmix_detection.txt	0.500342	Re-ranked DPM detection over Mixed selection from High-Level SVM scores and Baseline Scores, decision is performed by looking at the validation performance
OXFORD_VGG	test_finecls_detection_bestbbox.txt	0.50139	Re-ranked DPM detection over High-Level SVM Scores
OXFORD_VGG	test_finecls_detection_firstbbox.txt	0.522189	Re-ranked DPM detection over High-Level SVM Scores - First bbox selection heuristic

SuperVision (SV)

Image classification: Deep convolutional neural networks

- 7 hidden "weight" layers, 650K neurons, 60M parameters, 630M conn ections
- Rectified Linear Units, max pooling, dropout trick
- Randomly extracted 224x224 patches for more data
- Trained with SGD on two GPUs for a week, **fully supervised**

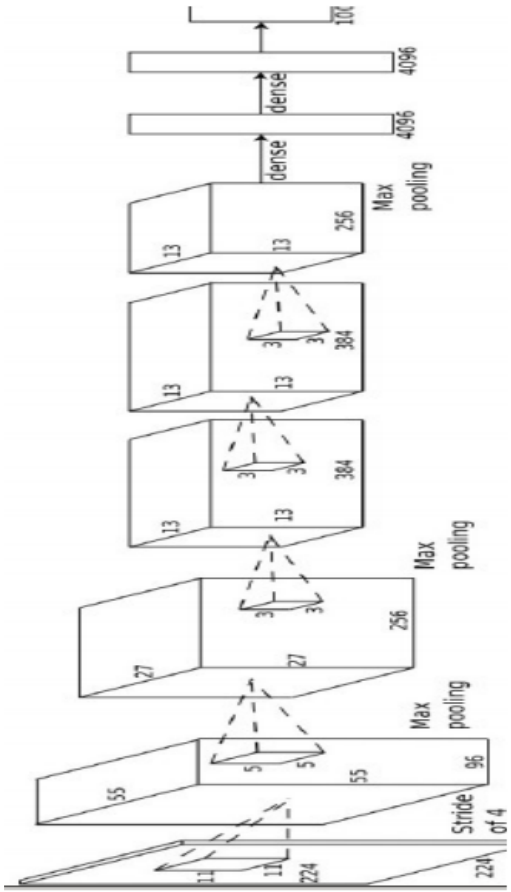
Localization: Regression on (x,y,w,h)

http://image-net.org/challenges/LSVRC/2012/supervision.pdf

SuperVision

- Won the 2012 ImageNet LSVRC. 60 Million parameters, 832M MAC ops

4M	FULL CONNECT
16M	FULL 4096/ReLU
37M	FULL 4096/ReLU
	MAX POOLING
442K	CONV 3x3/ReLU 256fm
1.3M	CONV 3x3ReLU 384fm
884K	CONV 3x3/ReLU 384fm
	MAX POOLING 2x2sub
	LOCAL CONTRAST NORM
307K	CONV 11x11/ReLU 256fm
	MAX POOL 2x2sub
	LOCAL CONTRAST NORM
35K	CONV 11x11/ReLU 96fm

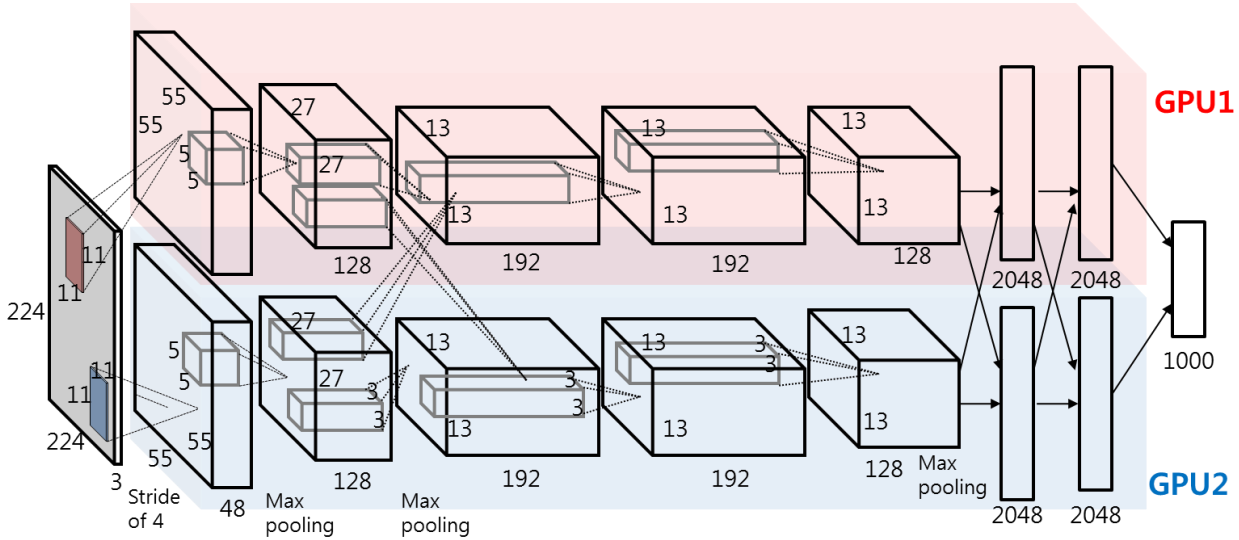


Object Recognition

ALEXNET

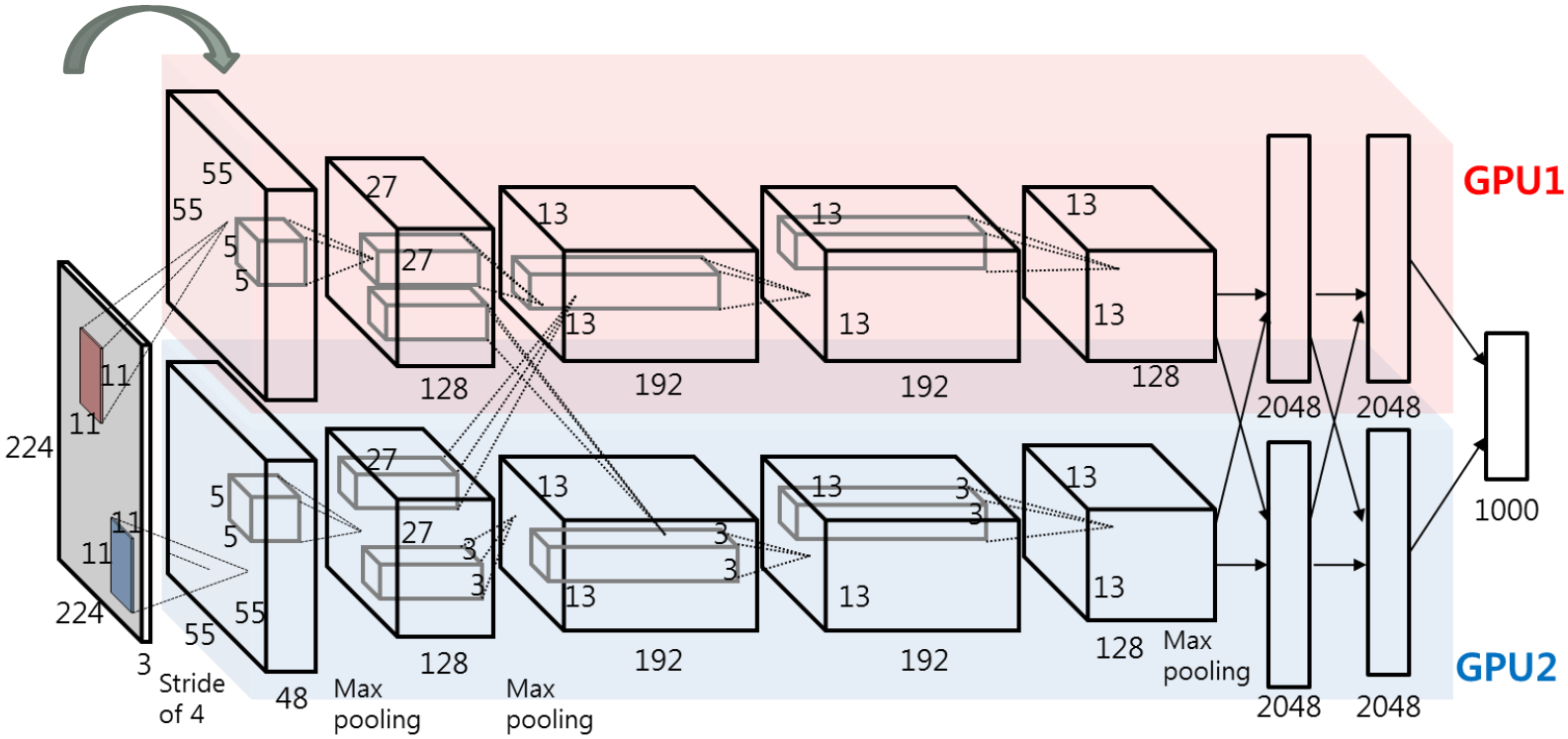
AlexNet

- AlexNet: won the 2012 ImageNet competition by making 40% less error than the next best competitor
  - It is composed of 5 convolutional layers
  - The input is a color RGB image
  - Computation is divided over 2 GPU architectures
  - Learning uses artificial data augmentation and connection drop-out to avoid over-fitting



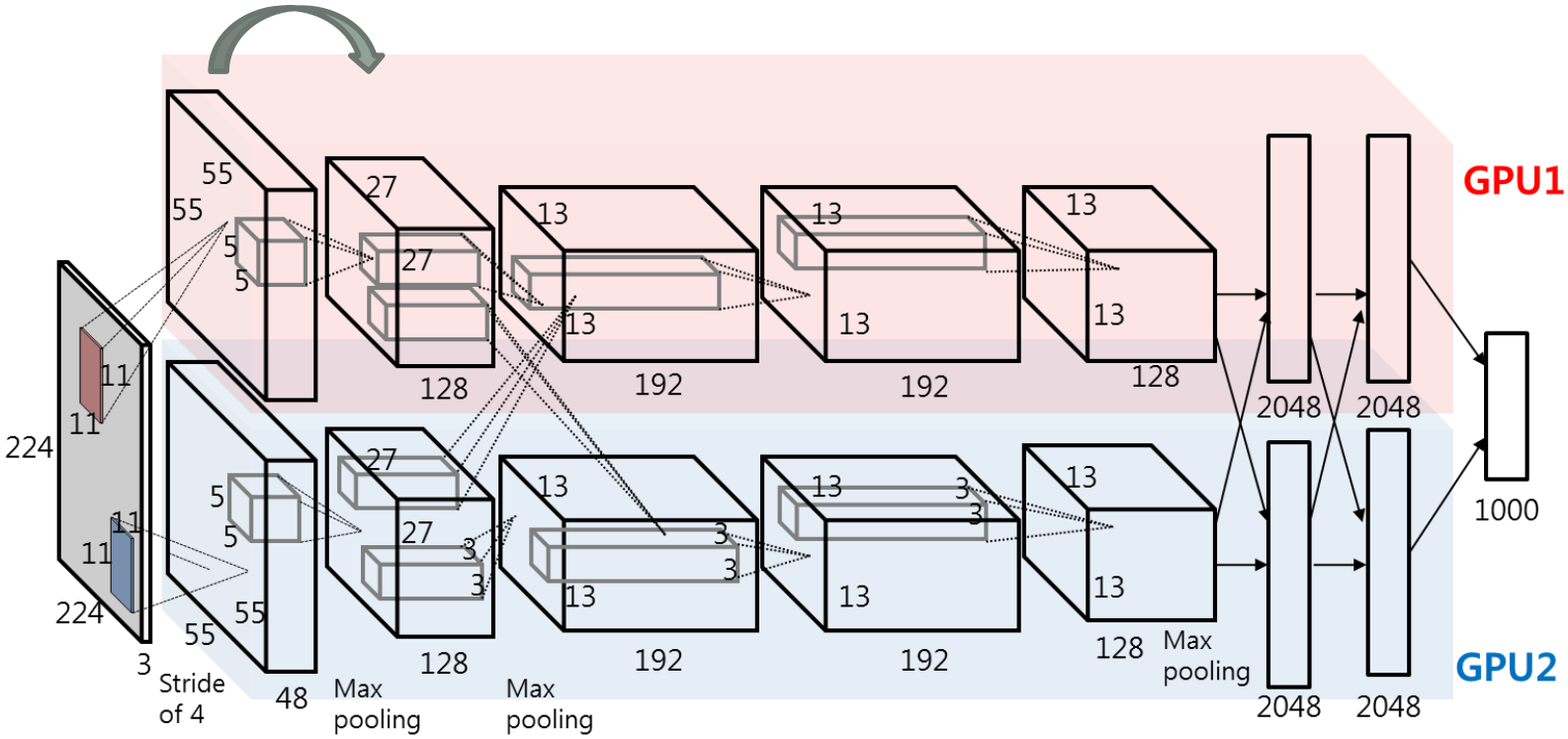
AlexNet in details

- The first layer applies 96 kernels of size 3x11x11
  - 34,848 parameters
  - Each kernel is applied with a stride of 4 pixels
  - (11x11x3)x(55x55x(48+48)) = 105,415,200 MACs



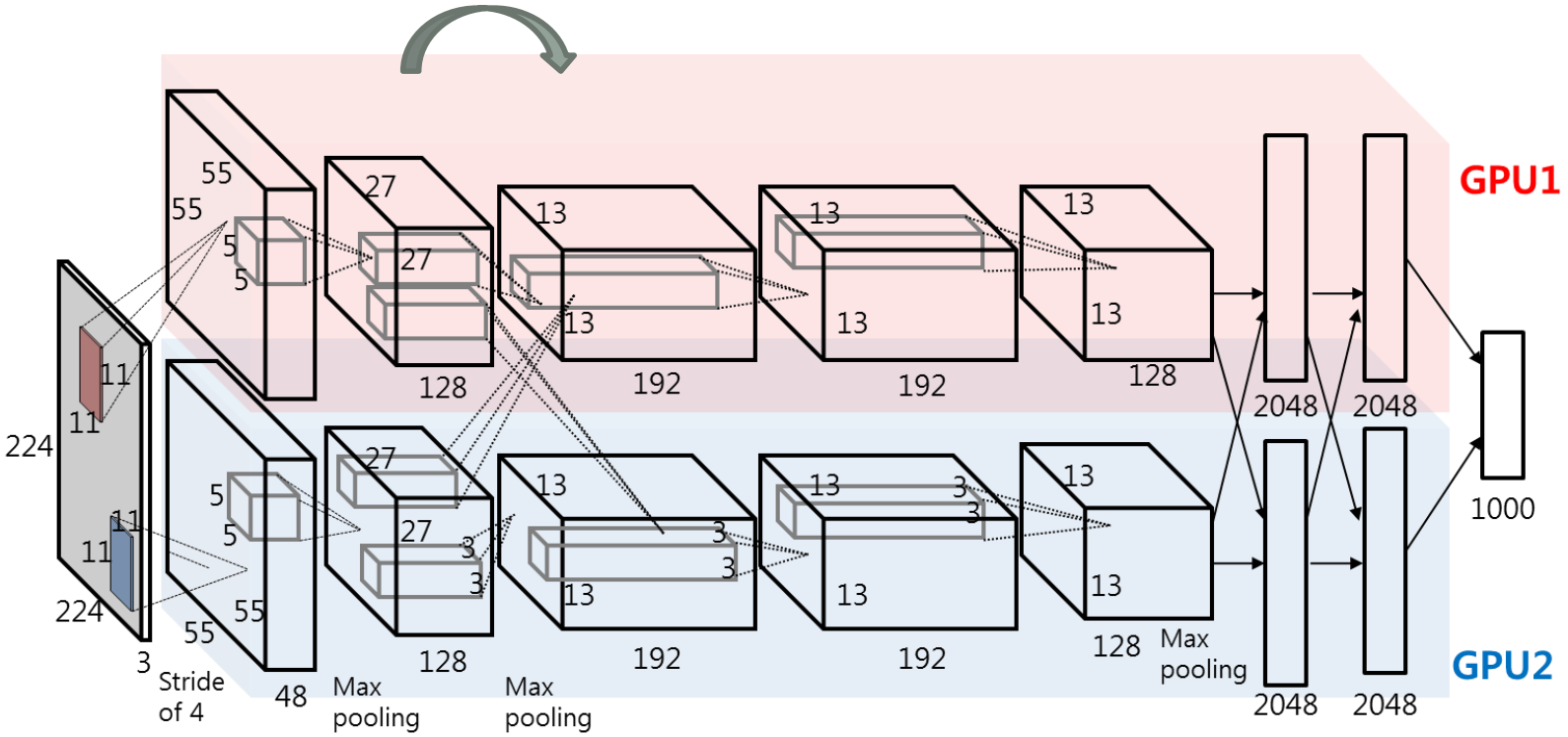
AlexNet in details

- The second layer applies 256 kernels of size 48x5x5
  - After applying a 3x3 max pooling with a stride of 2 pixels
  - 307,200 parameters
  - 256x(48x5x5)x(27x27)=223,948,800 MACs



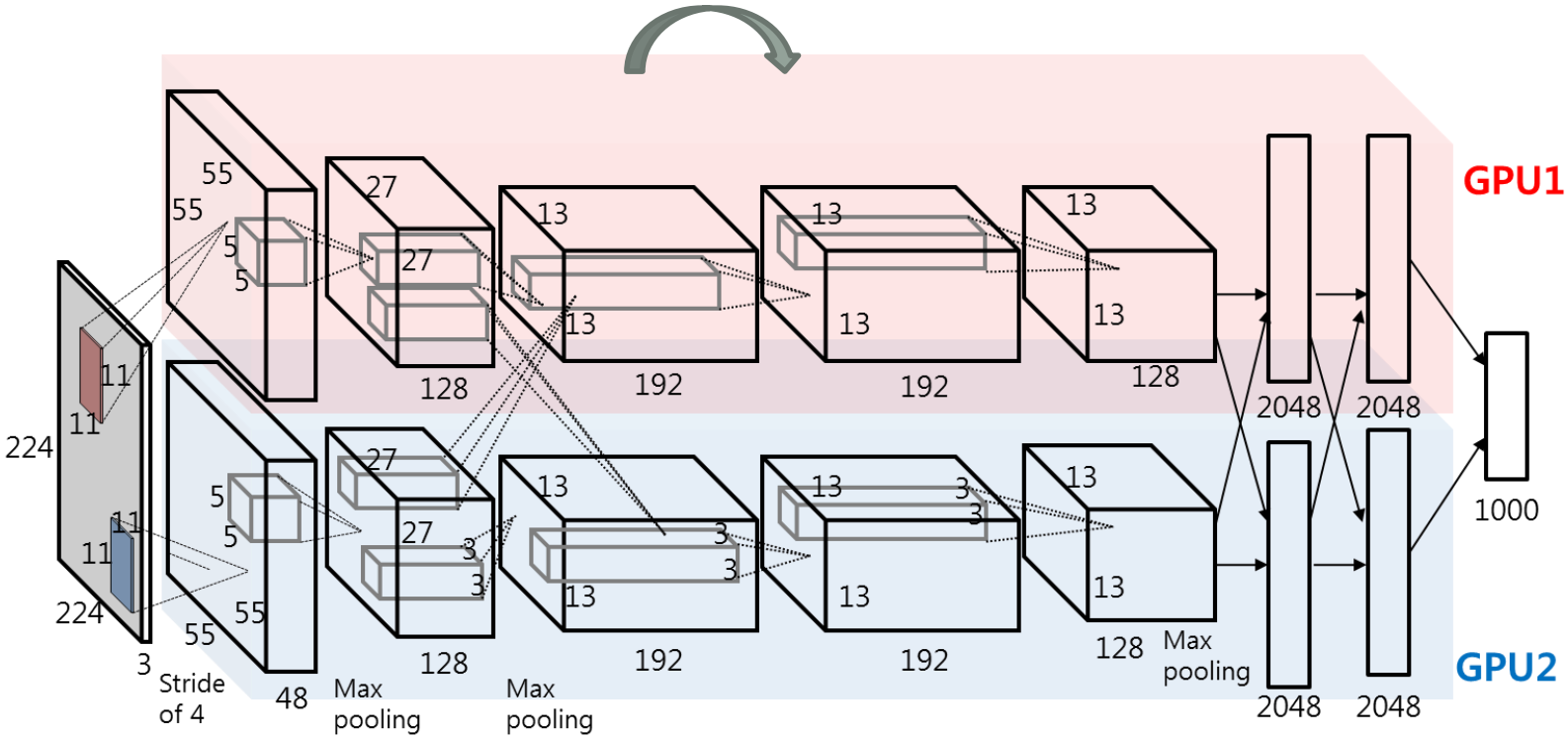
AlexNet in details

- The third layer applies 384 kernels of size 256x3x3
  - After applying a 3x3 max pooling with a stride of 2 pixels
  - 884,736 parameters
  - 384x((128+128)x3x3)x(13x13)=149,520,384 MACs



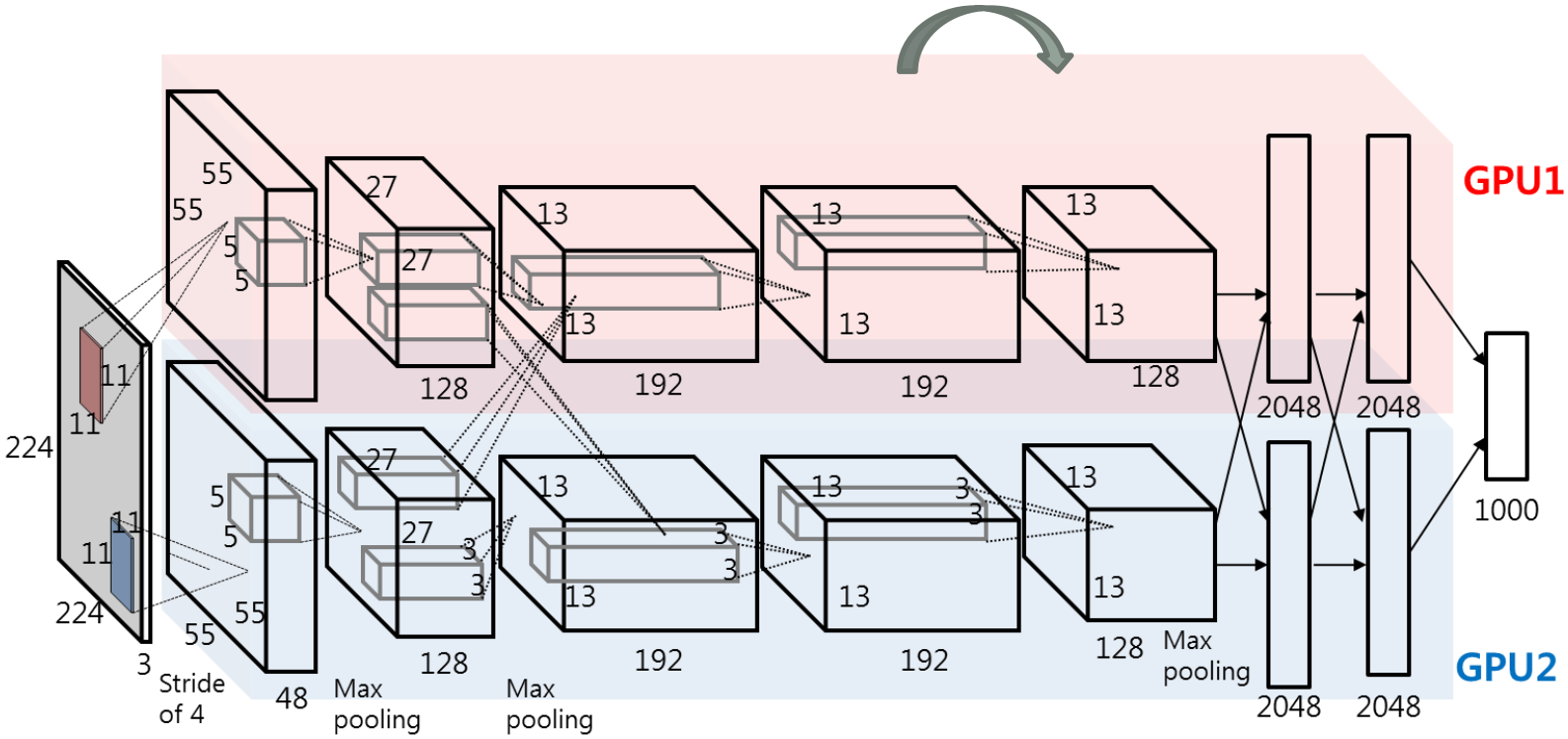
AlexNet in details

- The fourth layer applies 384 kernels of size 192x3x3
  - Without pooling
  - 663,552 parameters
  - 384x(192x3x3)x(13x13)=112,140,288 MACs



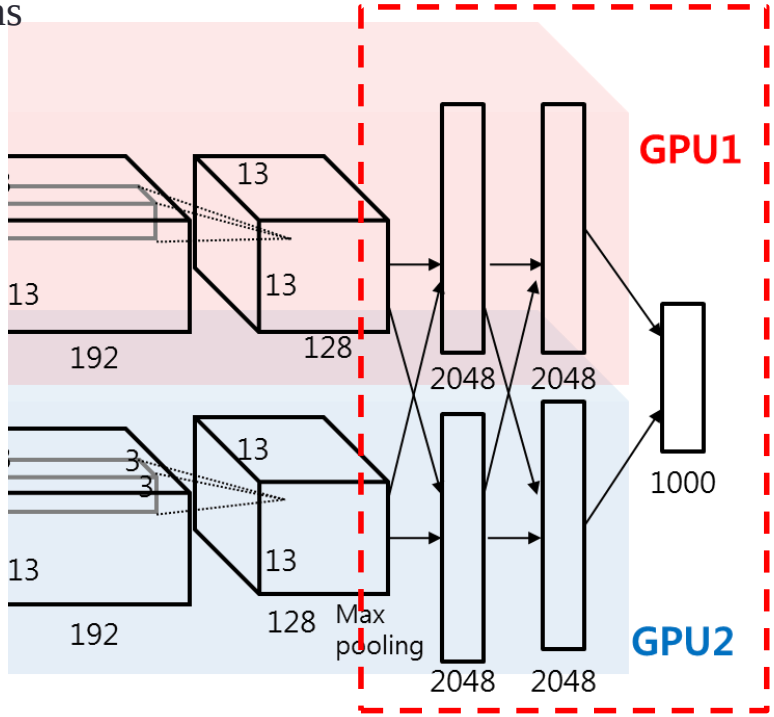
AlexNet in details

- The fifth layer applies 256 kernels of size 192x3x3
  - Without pooling
  - 442,368 parameters
  - 256x(192x3x3)x(13x13)=74,760,192 MACs



AlexNet in details

- The output of the fifth layer (after a 3x3 max pooling with a stride of 2 pixels) is connected to a fully connected 3-layer perceptron
  - 1st layer
    - (2x6x6x128)x4096= 37,748,736connections
  - 2nd layer
    - 4096x4096= 16,777,216 connections
  - 3rd layer
    - 4096x1000= 4,096,000 connections



AlexNet in details

- 60 Million parameters, 832M MAC ops