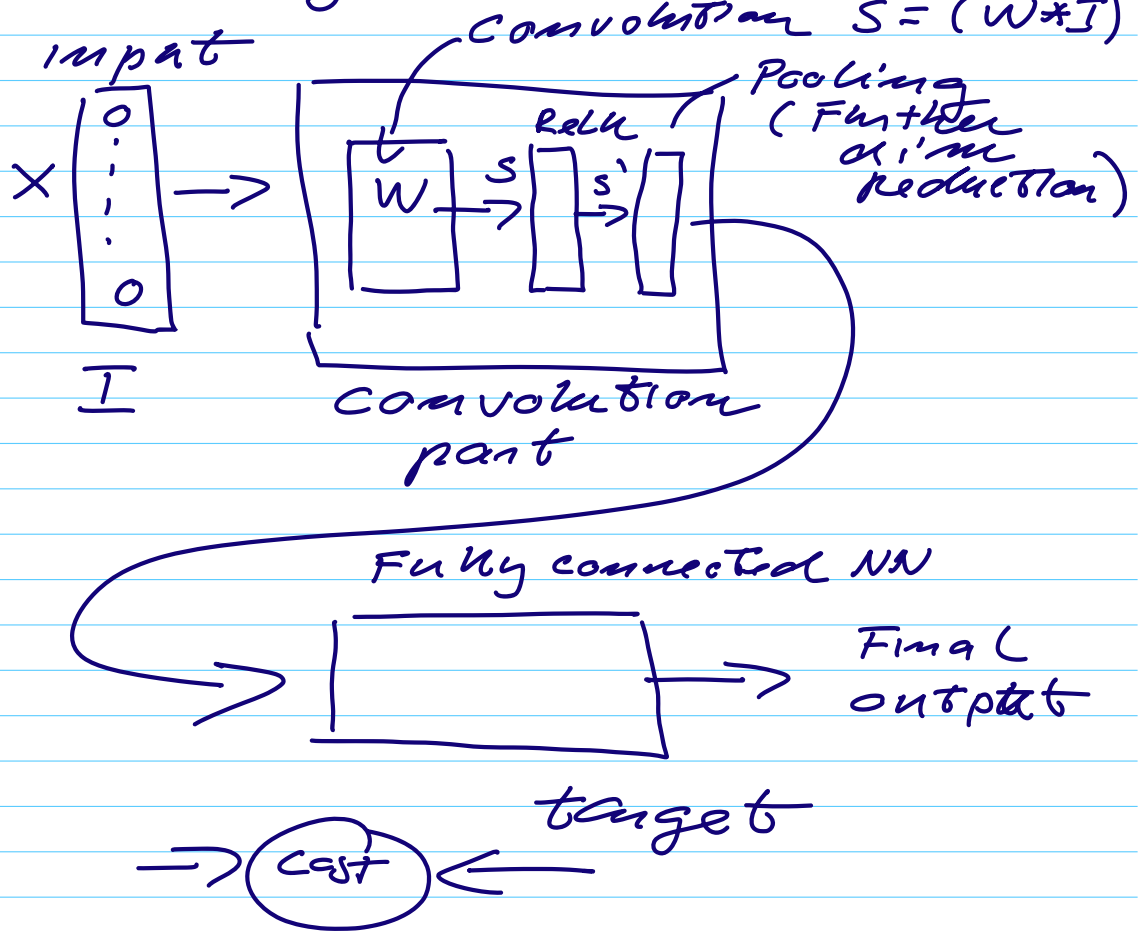
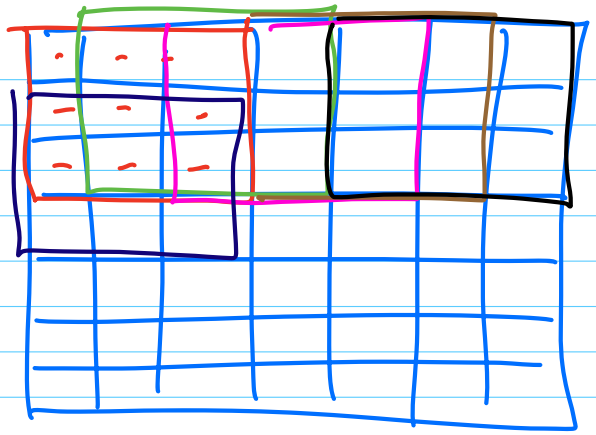
ECE 5429/6429, FEB 22, 2023

CNNs

Basic layout

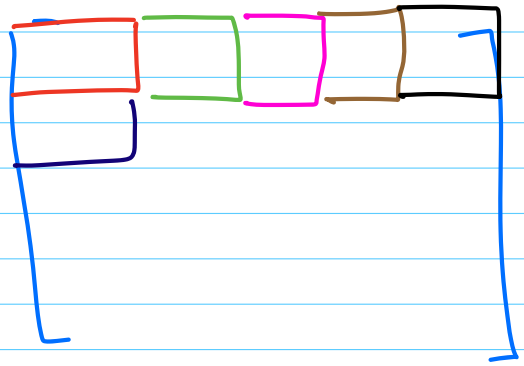


7x7 input image N = 7
3x3 Filter F = 3
Stride = 1 S = 1
padding = 0



3x3 Filter

S =



stride = 1

S is a 5 x 5 matrix

stride = 2

S is a 3 x 3 matrix

stride = 3

output size

$$(N - F)/S + 1$$

N = 7, F = 3, S = 1

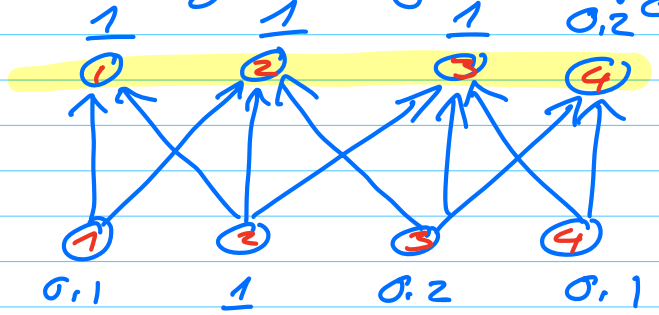
output is 5 => 5 x 5 matrix

$N = 7, \quad F = 3, \quad S = 2$

$(7-3)/2 + 1 = 3 \Rightarrow 3 \times 3$ matrix as output.

Typical convolution part

- one or several convolutions in parallel to produce a set of linear results which are fed into to the activation part
- each linear activation inputs are run through a non-linear activation function
- Pooling stage (sparse connectivity)



max pooling + down sampling.

Example

input volume $32 \times 32 \times 3$

10 5×5 filters with stride

= 1 and no padding

output volume

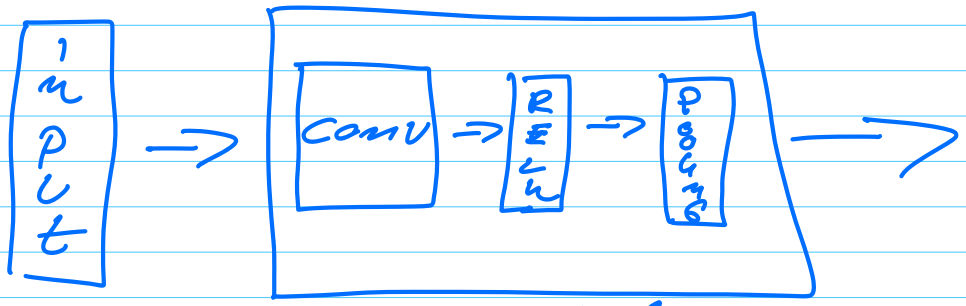
$(32-5)/1 + 1 = 28$

$28 \times 28 \times 10$ and include color channels = 3

Filter has $5 \times 5 \times 3 + 1$ (bias)

= 76 parameters

Ten images in total => 760 parameters to fit.



convolution part

new convolution part or FC = Fully connected NN.

in summary:

- accept a volume of $W_1 \times H_1 \times D_1$

width ↑ height ↑ depth ↑

- need four new hyper parameters
  - K = number of filters
  - F = their spatial extent
  - S = the stride
  - P = padding of zeros

produces an output

$$W_2 \times H_2 \times D_2$$

$$W_2 = (W_1 - F + 2P)/S + 1$$

$$H_2 = (H_1 - F + 2P)/S + 1$$

$$D_2 = K$$

We get then $F \times F \times D_1$ weights per filter for a total $(F \times F \times D_1) \times K$ weights and K biases (trained by Back propagation)

K is chosen in power of 2.

Common settings

$F = 3 \quad S = 1 \quad P = 1$

$F = 5 \quad S = 1 \quad P = 2$

$F = 5 \quad S = 2 \quad P = 0, 1, 2$