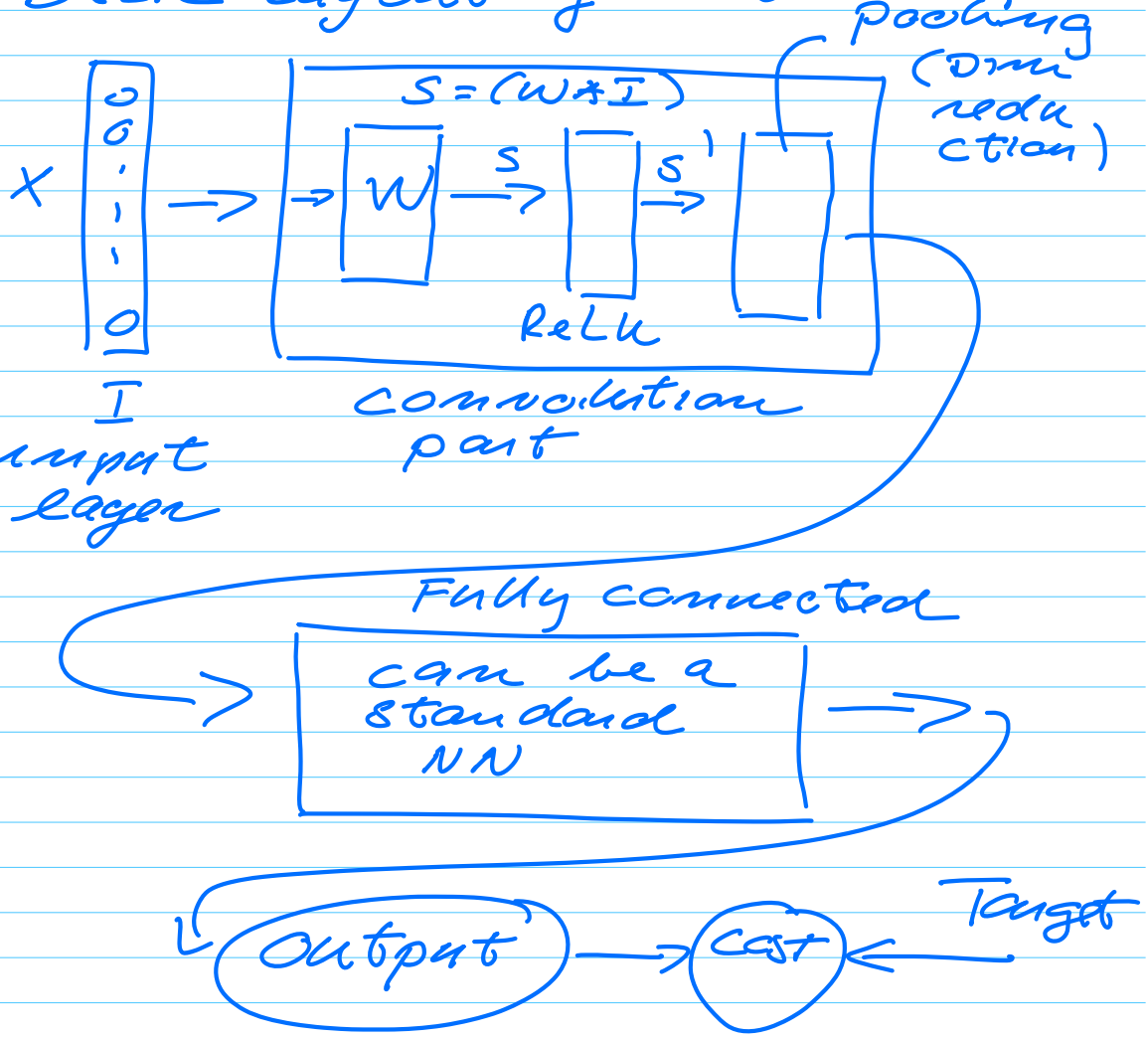
Comp Sci, FEB 28, 2023

CNNs

Basic layout of CNN



Two new hyperparameters

S = Stride

P = padding

F = dimensionality of Filter

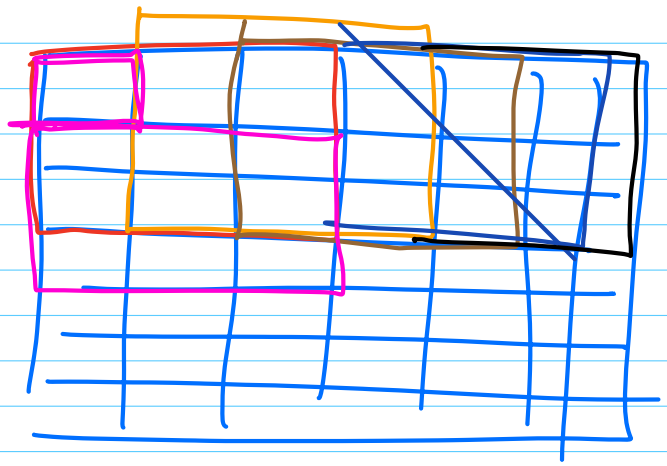
7 × 7 (× 1) input image

$N = 7$

3 × 3 (× 1) Filter $F = 3$

Stride $S = 1$

padding $P = 0$



3 × 3

S is the output matrix

5 × 5 with stride = 1

S =

Stride $S = 2$ leads the matrix $S \in \mathbb{R}^{3 \times 3}$

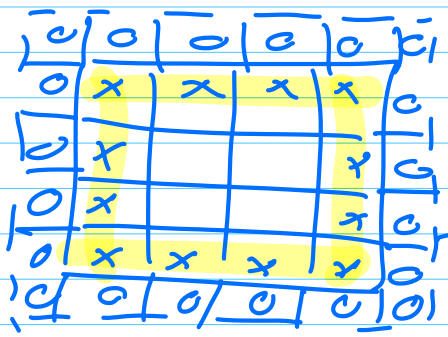
output size

$(N - F)/S + 1$

$N = 7 \quad F = 3 \quad S = 1 \Rightarrow Dim = 5 \times 5$

$N = 7 \quad F = 3 \quad S = 2 \Rightarrow Dim = 3 \times 3$

Padding $P = 0$ new hyperparameter.

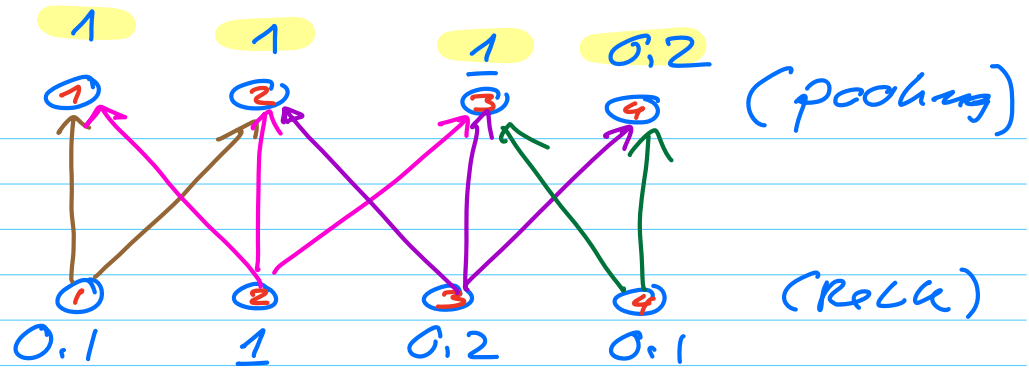


no padding

$P = 1$

Typical convolution part

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



max pooling + down sampling -

Example

input value $32 \times 32 \times 3$

10 outputs from convolution 5×5 filters with stride $S = 1$ and no padding

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

$28 \times 28 \times 10$ images + 3 color channels

Filter has $5 \times 5 \times 3 + 1$

$= 76$ parameters

Ten images $\Rightarrow$ 760 parameters to fit.

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 and K-biases (trained by back propagation)

K is often chosen in powers of 2.

Common settings

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

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

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