1. Convolution

$$\begin{bmatrix} 3 & 0 & 1 & 2 \\ 1 & 5 & 8 & 9 \\ 2 & 7 & 2 & 5 \\ 0 & 1 & 3 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

↑ filter

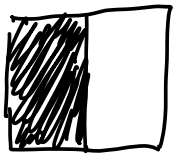
$$A = \begin{bmatrix} 3 & 0 & 1 \\ 1 & 5 & 8 \\ 2 & 7 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}$$

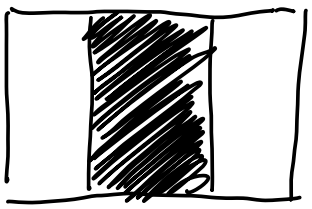
$$= 3\times1 + 0\times0 + 1\times-1 + 1\times1 + 5\times0 + 8\times-1 + 2\times1 + 7\times0 + 2\times-1$$

$$= -5$$

Intuition : Edge detection

$$\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 0 & -3 & -3 & 0 \\ 0 & -3 & -3 & 0 \\ 0 & -3 & -3 & 0 \\ 0 & -3 & -3 & 0 \end{bmatrix}$$





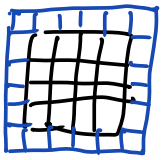
① Object A v. Object B ⇒ edge (A>B)

② vertical edge

③ $n\times n * f\times f = n-f+1 \times n-f+1$

2. Padding.

{ Valid conv : p = 0
  Same conv : input size = output size



4x4

5x5 p=1

$$\begin{bmatrix} 3 & 0 & 1 & 2 \\ 1 & 5 & 8 & 9 \\ 2 & 7 & 2 & 5 \\ 0 & 1 & 3 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

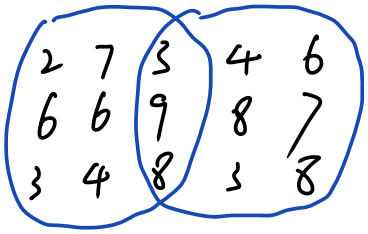
↑ filter

0 0 0 0 0 0
0 3 0 1 2 0
0 1 5 8 9 0
0 2 7 2 5 0
0 0 1 3 1 0
0 0 0 0 0 0

Intuition : ① image shrinks
② margin only used once

→ Output : $n+2p-f+1 \times n+2p-f+1$

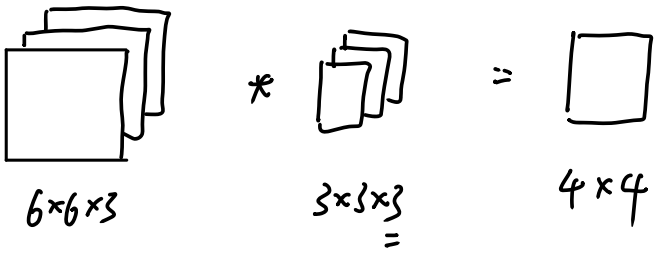
3. Stride. Move by more than 1 step.



Stride = 2

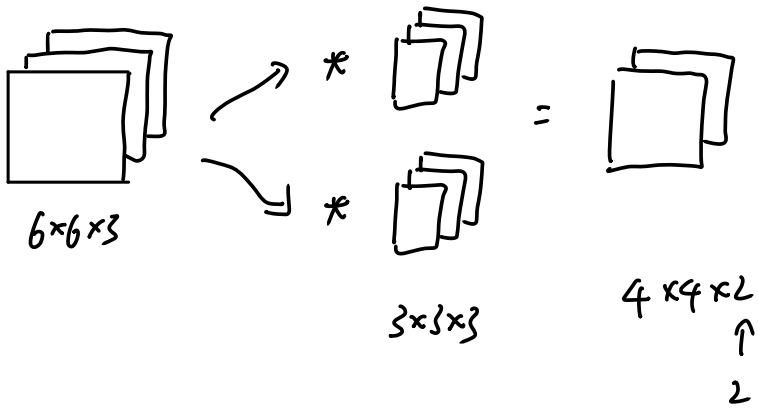
$$\Rightarrow \text{output: } \left\lfloor \frac{n+2p-f}{s} \right\rfloor + 1 \times \left\lfloor \frac{n+2p-f}{s} \right\rfloor + 1$$

4. Convolution on volume



RGB

5. Apply multiple filters



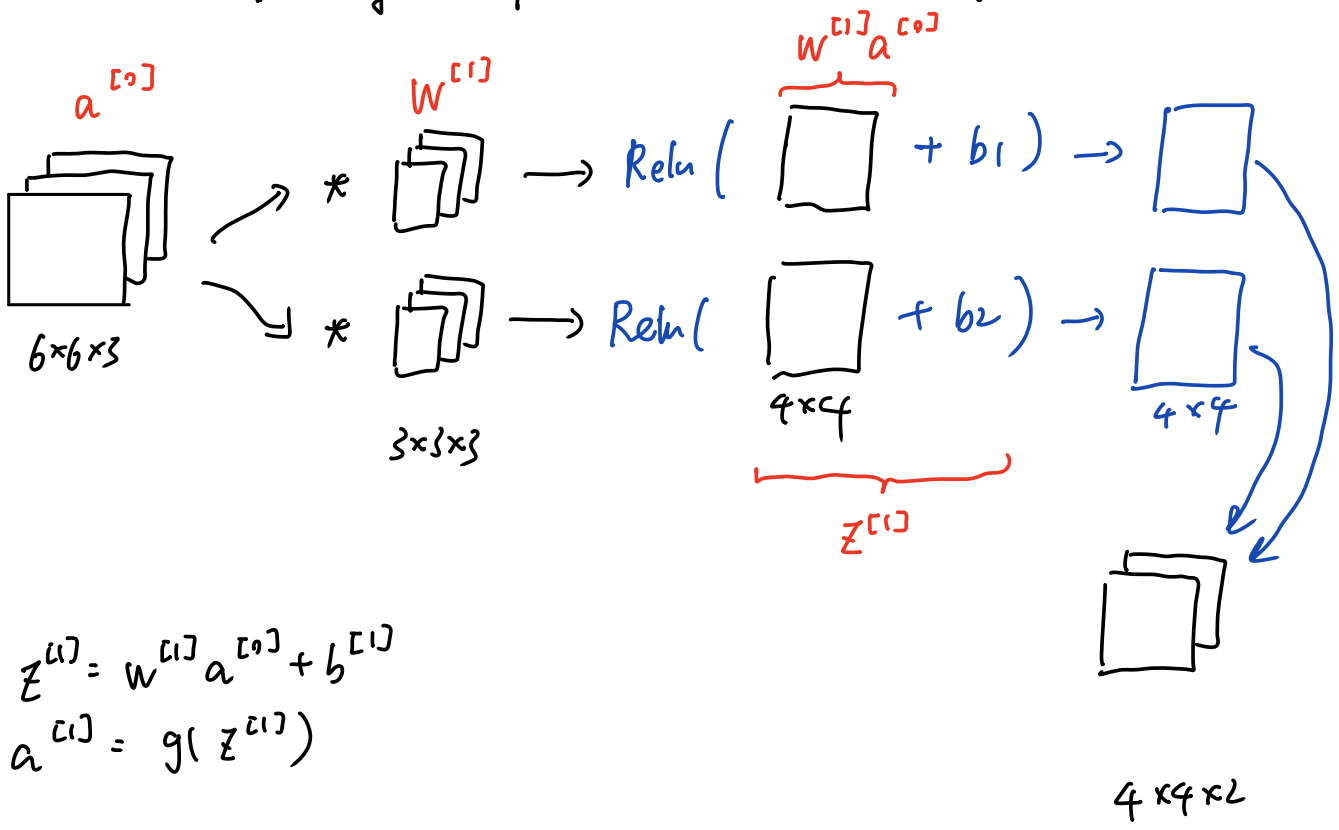
Recap: Convolution $1 \to N_c$ Number of channels

↓

padding

stride

6 Toy example of CNN Conv Layer



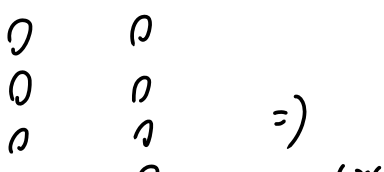
$$Z^{[1]} = W^{[1]} a^{[0]} + b^{[1]}$$

$$a^{[1]} = g(Z^{[1]})$$

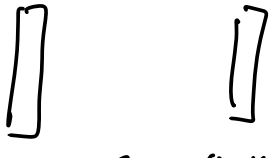
① 3-dim. Not flat

② 3×3×3 filter → 27 params, 10 filters ⇒ 270 params + bias ⇒ 280 params

Recall:



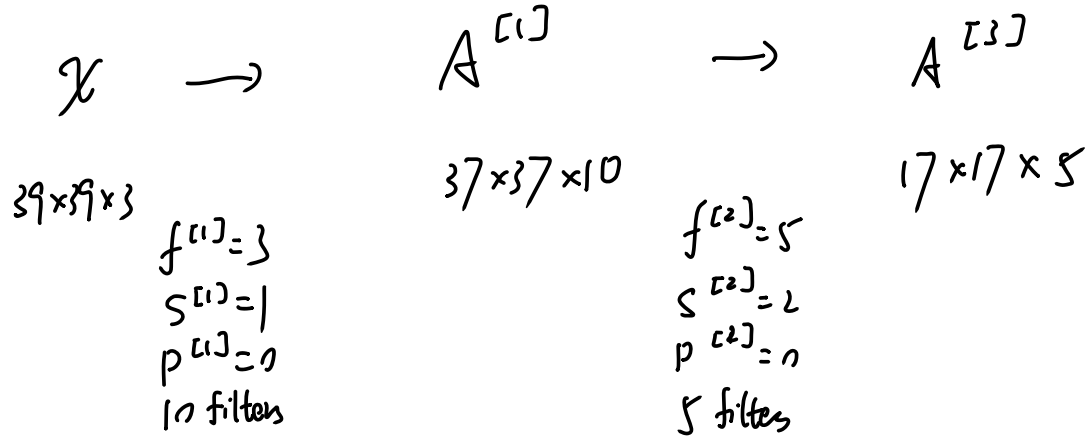
⇒



6×6×3 = 108

4×4×2 = 32

108×32 ≈ 3000



≈ 4800

≈ 2000

$$n^{[l]} = \left\lfloor \frac{n^{[l-1]} + 2p^{[l]} - f^{[l]}}{s^{[l]}} \right\rfloor + 1$$

Recap: Convolution $1 \to N_c$ Number of channels

↓

padding

stride

7. Types of Layer in CNN

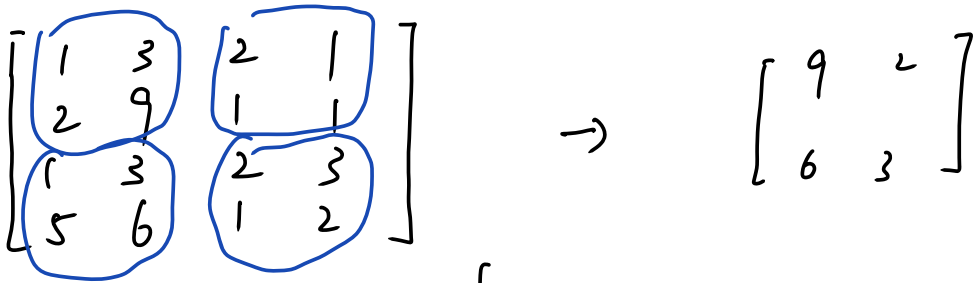
Convolution

Pooling

Fully Connected (Dense)

$X \to conv1 \to pool1 \to conv2 \to pool2 \to FC1 \to FC2 \to FC3 \to \hat{y}$ v. y

8. Pooling : Max pooling

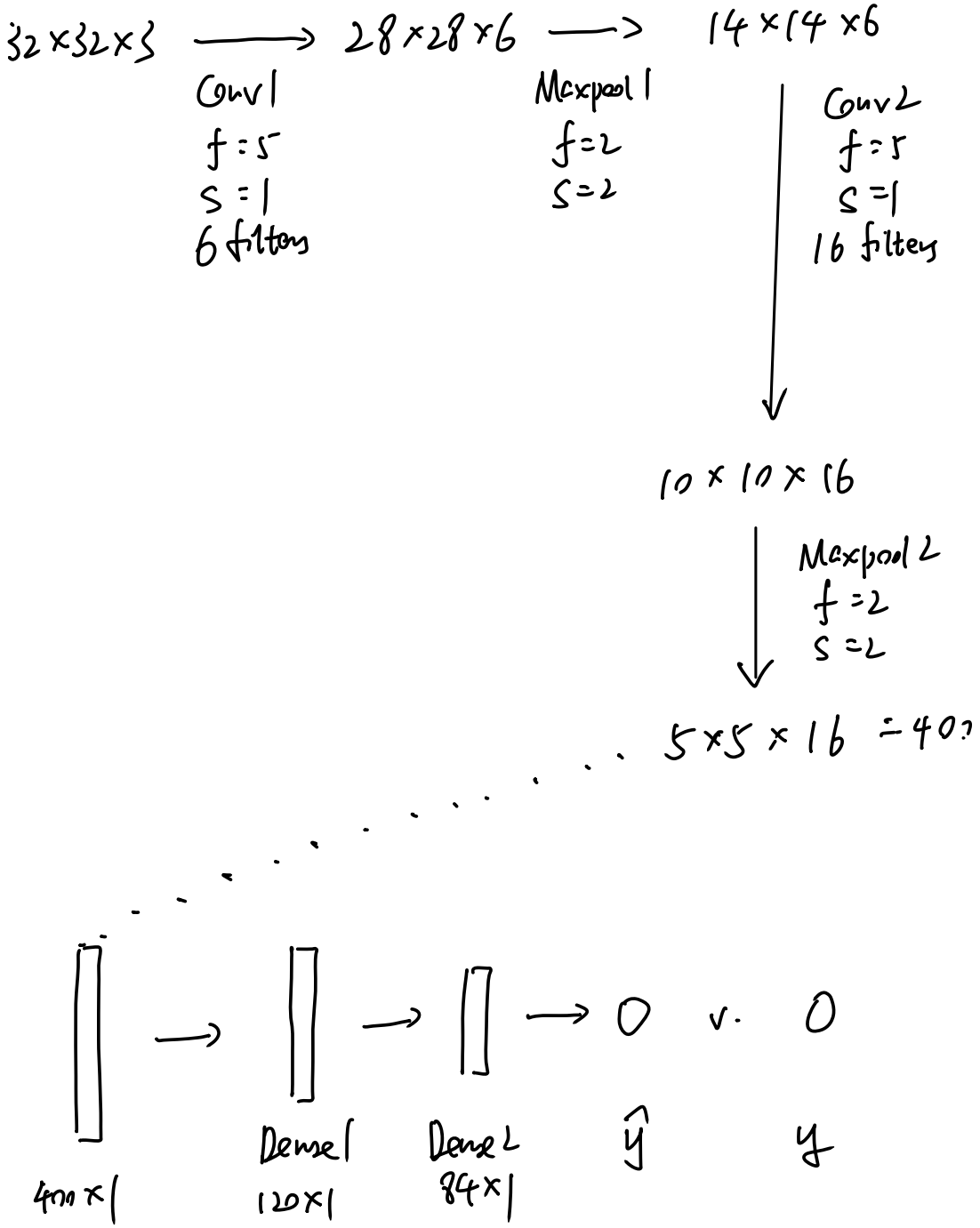


$f=2,\ s=2$

① speed up

② No params to Learn

9. Full example



Recap:

1. Conv → padding, stride
   Pooling.

2. X → conv1 → pool1 → conv2 → pool2 → FC1 → FC2 → FC3 → $\hat{y}$ v. y

3. Why?
   (1) parameter sharing

$32 \times 32 \times 3 \longrightarrow 28 \times 28 \times 6$

Conv1
f = 5
s = 1
6 filters