# CONVOLUTIONAL NEURAL NETWORKS

## CONVOLUTION OPERATION

#STEP 1

$$3\cdot1+0\cdot0+1\cdot(-1)+1\cdot1+5\cdot0+8\cdot(-1)+2\cdot1+7\cdot0+2\cdot(-1)$$

| | | | | | |
|---|---|---|---|---|---|
| 3 (1) | 0 (0) | 1 (-1) | 2 | 7 | 4 |
| 1 (1) | 5 (0) | 8 (-1) | 9 | 3 | 1 |
| 2 (1) | 7 (0) | 2 (-1) | 5 | 1 | 3 |
| 0 | 1 | 3 | 1 | 7 | 8 |
| 4 | 2 | 1 | 6 | 2 | 8 |
| 2 | 4 | 5 | 2 | 3 | 9 |

6×6

*

Filter

| | | |
|---|---|---|
| 1 | 0 | -1 |
| 1 | 0 | -1 |
| 1 | 0 | -1 |

3×3

=

| | | | |
|---|---|---|---|
| -5 | | | |
| | | | |
| | | | |
| | | | |

4×4

"Cross correlation" by convention in ML is called "Convolution"

#STEP 2

| | | | | | |
|---|---|---|---|---|---|
| | 0 (1) | 1 (0) | 2 (-1) | | |
| | 5 (1) | 8 (0) | 9 (-1) | | |
| | 7 (1) | 2 (0) | 5 (-1) | | |
| | | | | | |
| | | | | | |
| | | | | | |

*

Filter

| | | |
|---|---|---|
| 1 | 0 | -1 |
| 1 | 0 | -1 |
| 1 | 0 | -1 |

3×3

=

| | | | |
|---|---|---|---|
| -5 | -4 | | |
| | | | |
| | | | |
| | | | |

4×4

## VERTICAL EDGE DETECTION

We want bright pixels (left) & dark pixels (right)

| | | | | | |
|---|---|---|---|---|---|
| 10 | 10 | 10 | 0 | 0 | 0 |
| 10 | 10 | 10 | 0 | 0 | 0 |
| 10 | 10 | 10 | 0 | 0 | 0 |
| 10 | 10 | 10 | 0 | 0 | 0 |
| 10 | 10 | 10 | 0 | 0 | 0 |
| 10 | 10 | 10 | 0 | 0 | 0 |

*

| | | |
|---|---|---|
| 1 | 0 | -1 |
| 1 | 0 | -1 |
| 1 | 0 | -1 |

3×3

=

| | | | |
|---|---|---|---|
| 0 | 30 | 30 | 0 |
| 0 | 30 | 30 | 0 |
| 0 | 30 | 30 | 0 |
| 0 | 30 | 30 | 0 |

4×4

EDGE DETECTED

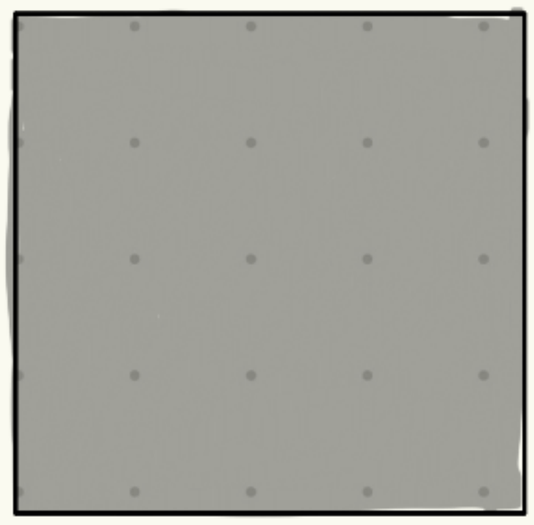
White line is bigger because input img is small (6×6)

## HORIZONTAL EDGE DETECTION

From Bright to Dark

$$\begin{bmatrix} 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 10 & 10 & 10 & 0 & 0 & 0 \\ 0 & 0 & 0 & 10 & 10 & 10 \\ 0 & 0 & 0 & 10 & 10 & 10 \\ 0 & 0 & 0 & 10 & 10 & 10 \end{bmatrix} * \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 30 & 10 & -10 & -30 \\ 30 & 10 & -10 & -30 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

## LEARNING TO DETECT EDGES

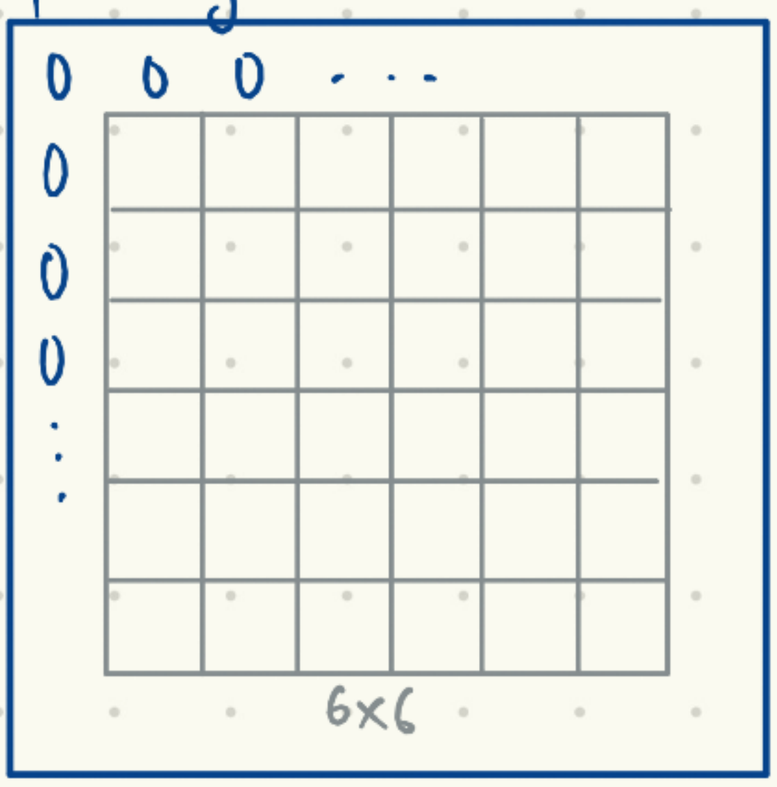


$$* \underbrace{\begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix}}_{\text{Learning filters}} = \begin{bmatrix} \quad \end{bmatrix}$$

## PADDING

Downsides to solve: → Shrinking img when using filters
→ Losing info from the edge

padding = 1




$$* \underbrace{\square}_{3\times3} = \underbrace{\square}_{6\times6}$$

$$\cancel{n}+2p-f+1=\cancel{n}$$

$$\boxed{p=\frac{f-1}{2}}$$

$$D = n-f+1 \begin{cases} \rightarrow \text{Img } 6\times6,\ f=3\times3 \rightarrow D=6-3+1=4 \\ \rightarrow \text{Img } 8\times8,\ f=3\times3 \rightarrow D=8-3+1=6 \end{cases}$$

## STRIDE

Stride → Saltos

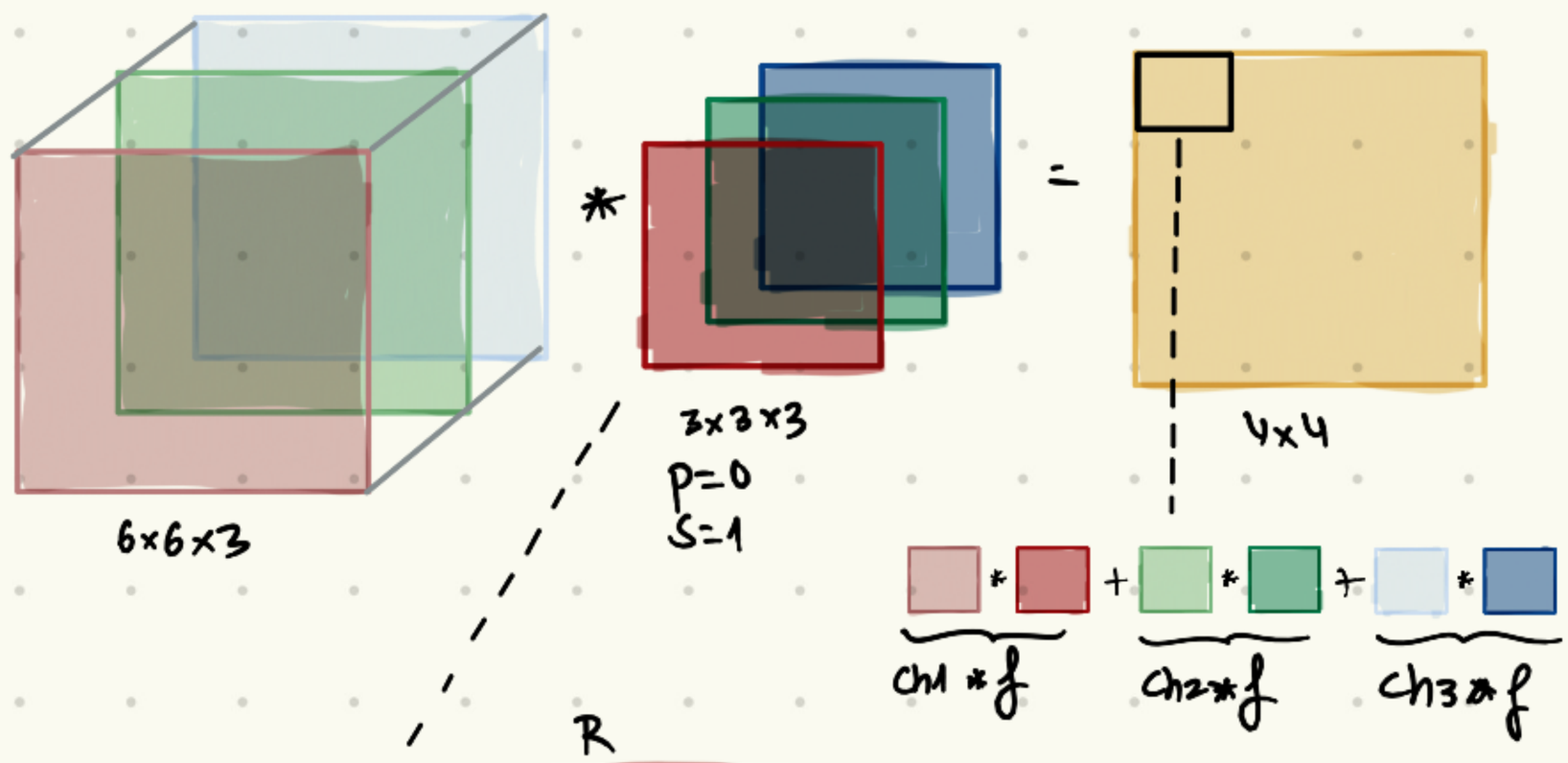
$$\boxed{D = \left\lfloor \frac{n+2p-f}{s} + 1 \right\rfloor}$$

E.g.: $n=7$, $p=0$, $f=3$, $s=2$

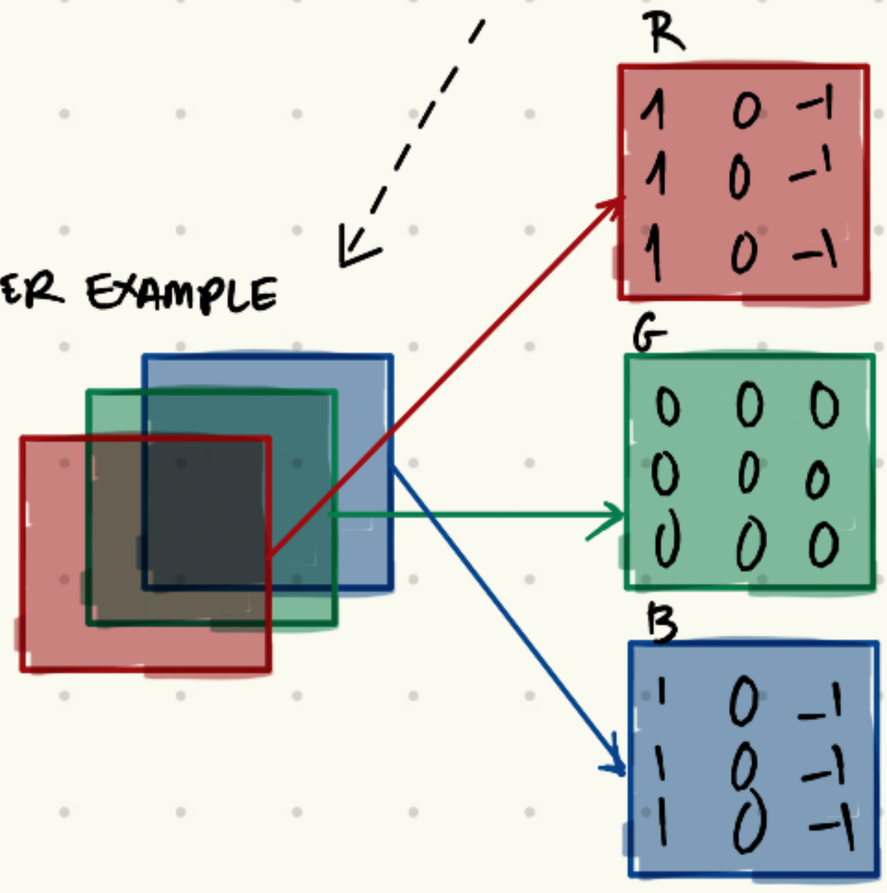
$$D = \frac{7-3}{2} + 1 = 3$$

## CONVOLUTION ON RGB IMGS

MISMO FILTRO

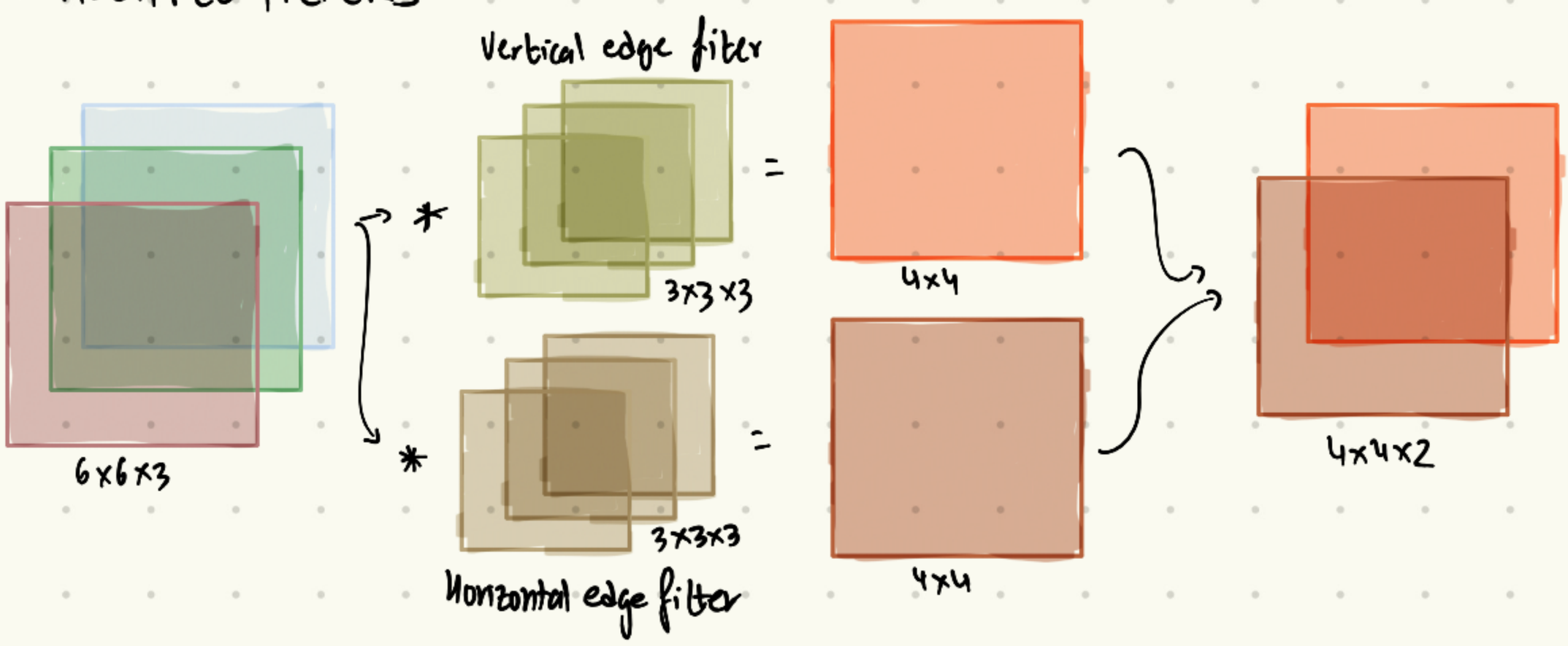



FILTER EXAMPLE



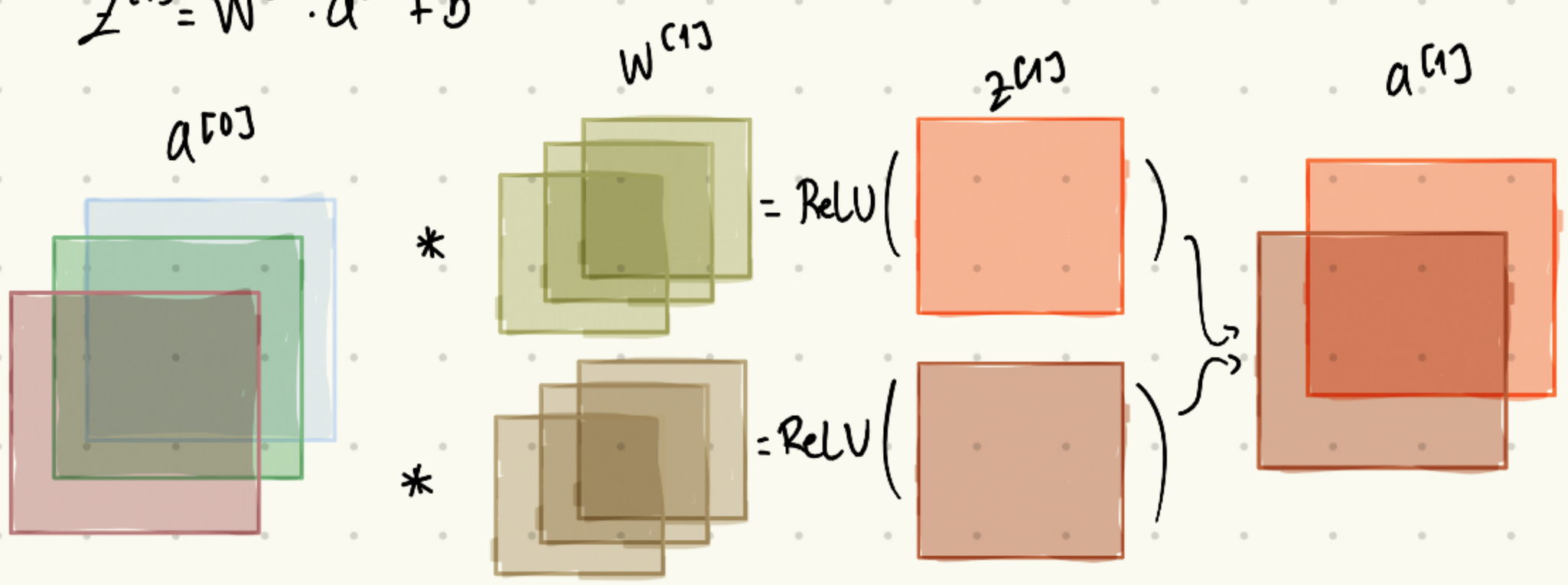

Detect vertical edges on Red & Blue channels

▶ MULTIPLE FILTERS



## ONE LAYER OF A CNN

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




¿NUMBER OF PARAMETER?

Example: 10 filters, 3x3x3. Nº parameters in one layer


$+ \underset{1\,par}{b} = 28$ param/filter $\rightarrow$ Total $= 28 \cdot 10 = \underline{280 \text{ param}}$

## NOTATION SUMMARY

In layer "$l$":

- $f^{[l]}$ = filter size
- $p^{[l]}$ = padding size
- $s^{[l]}$ = stride size
- $n_c^{[l]}$ = number of filters/channels
- Each filter is: $f^{[l]} \times f^{[l]} \times n_c^{[l-1]}$
- Activations: $a^{[l]} \rightarrow n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}$
- Weights: $\underbrace{f^{[l]} \times f^{[l]} \times n_c^{[l-1]}}_{\text{one filter}} \times n_c^{[l]}$
- Bias: $n_c^{[l]}$
  $\hookrightarrow (1, 1, 1, n_c^{[l]})$

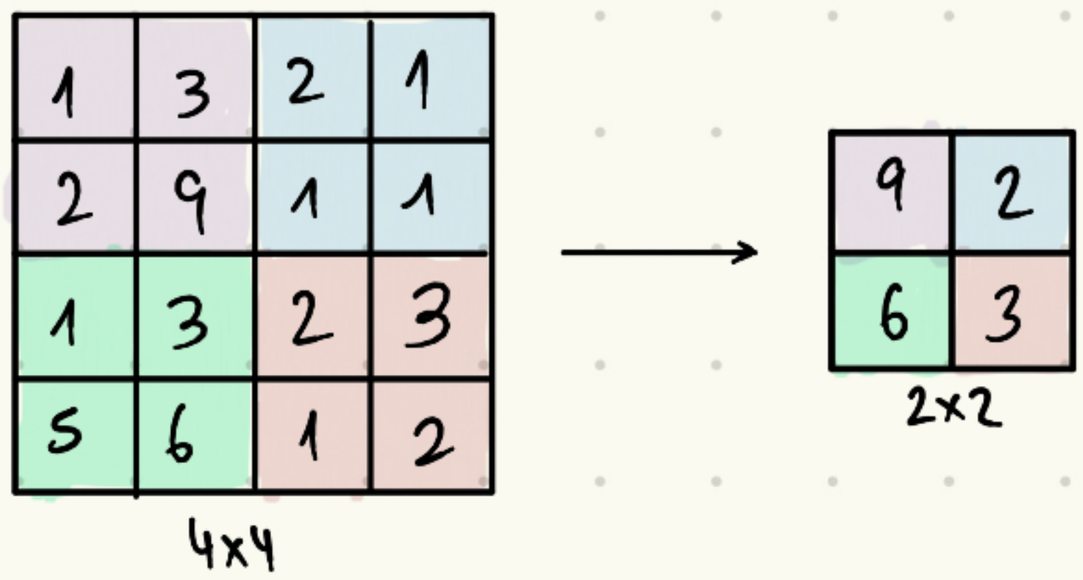
- INPUT: $n_H^{[l-1]} \times n_W^{[l-1]} \times n_c^{[l-1]}$
- OUTPUT: $n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}$

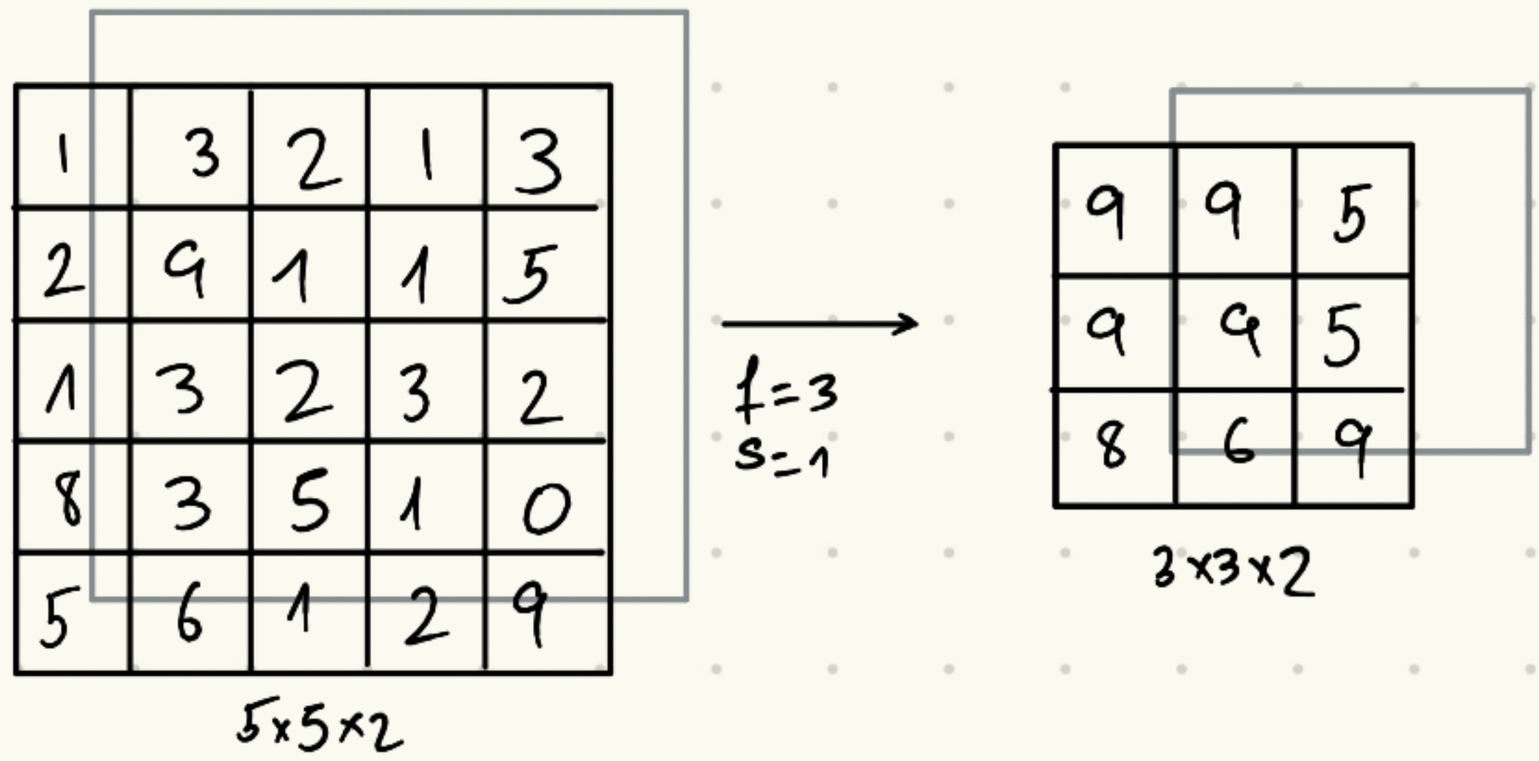
$$\rightarrow n_H^{[l]} = \frac{n_H^{[l-1]} + 2p^{[l]} - f^{[l]}}{s^{[l]}} + 1$$

$$\rightarrow n_W^{[l]} = \frac{n_W^{[l-1]} + 2p^{[l]} - f^{[l]}}{s^{[l]}} + 1$$

$$A^{[l]} = m \times n_H^{[l]} \times n_W^{[l]} \times n_c^{[l]}$$

## MAX-POOLING

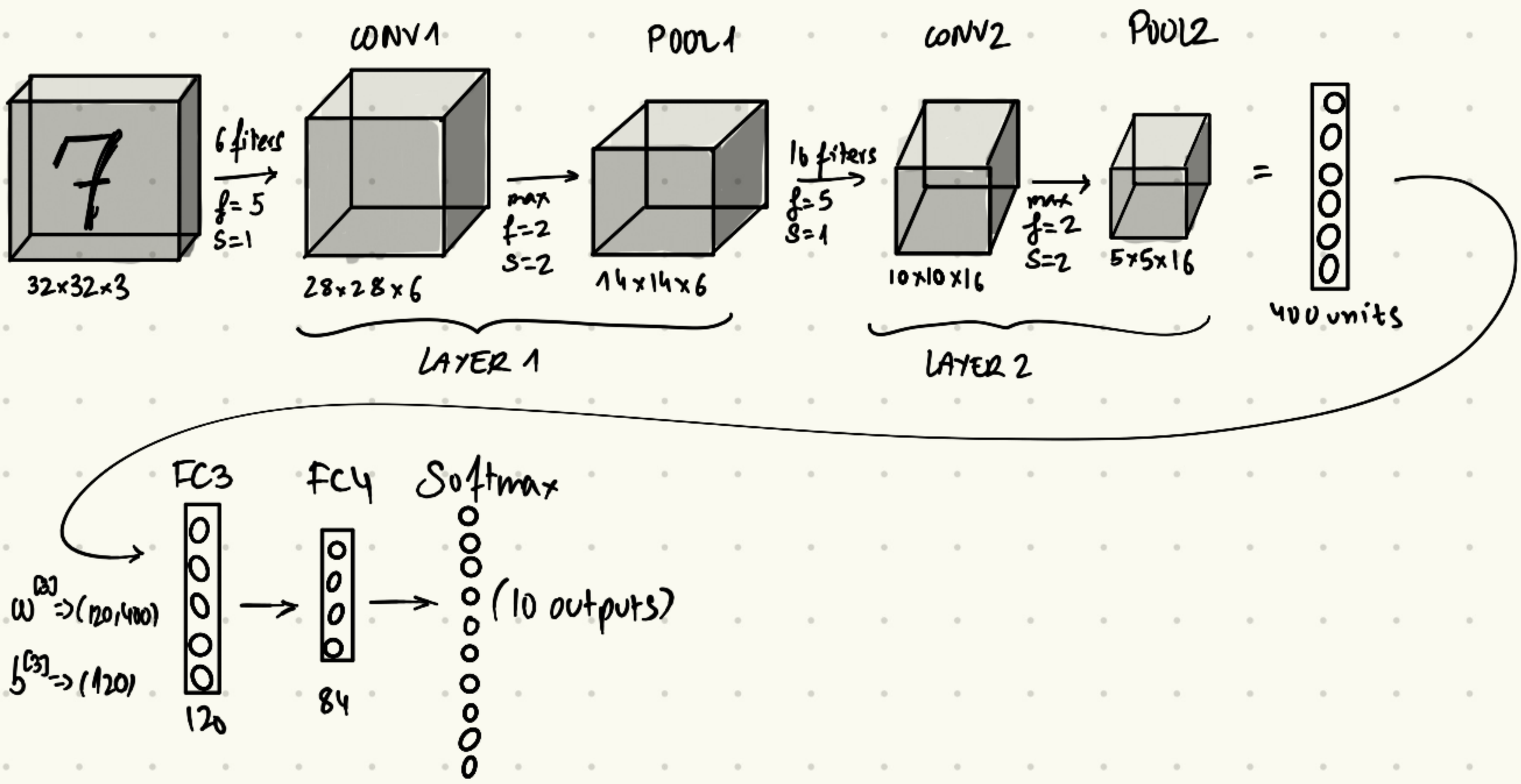



## AVERAGE POOLING

Similar to max-pooling but calculating the average

## LeNet-5



| | Activation shape | Activation size | # parameters |
|---|---|---|---|
| Input | (32,32,3) | 3072 | 0 |
| Conv1 (f=5, s=1) | (28,28,6) | 4707 | $3 \times 5 \times 5 \times 6 + 6 = 456$ |
| Pool1 | (14,14,6) | 1176 | 0 |
| Conv2 (f=5, s=1) | (10,10,16) | 1600 | 2416 |
| Pool2 | (5,5,16) | 400 | 0 |
| FC3 | (120,1) | 120 | 48120 |
| FC4 | (84,1) | 84 | 10164 |
| Softmax | (10,1) | 10 | 850 |