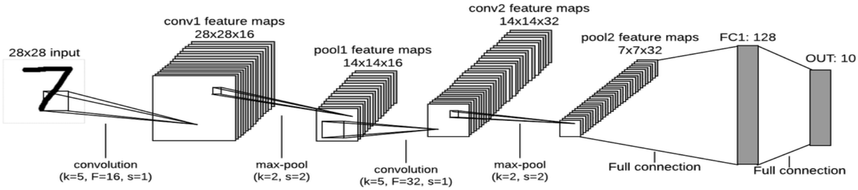
Redes convolucionales

nociones básica para se aplicadas

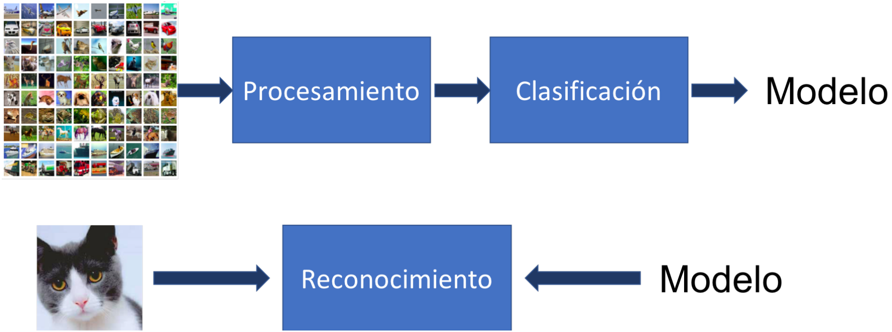
Julio Waissman



Plan de la presentación

1. Información local vs información global
2. ¿Cómo son las redes neuronales convolucionales?
3. ¿Cómo funciona esto?
4. Consideraciones finales

¿Qué es el reconocimiento de imágenes



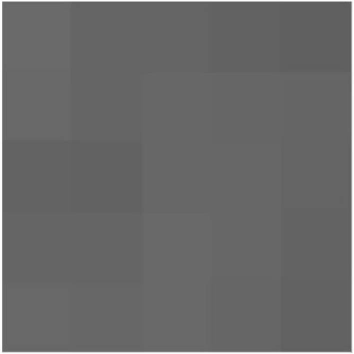
Imagen

Imagen original

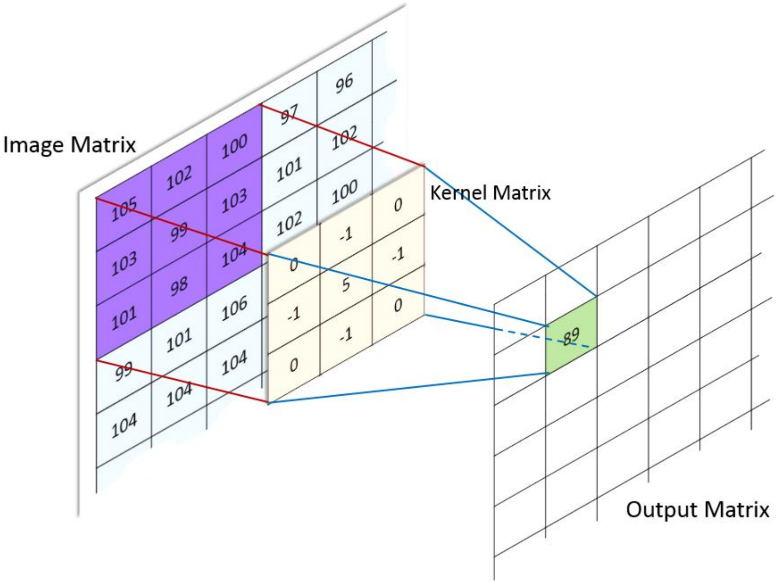
Imagen como matriz de números

Detalle

105	102	100	97	96
103	99	103	101	102
101	98	104	102	100
99	101	106	104	99
104	104	104	100	98



Filtro convolucional



Filtro convolucional

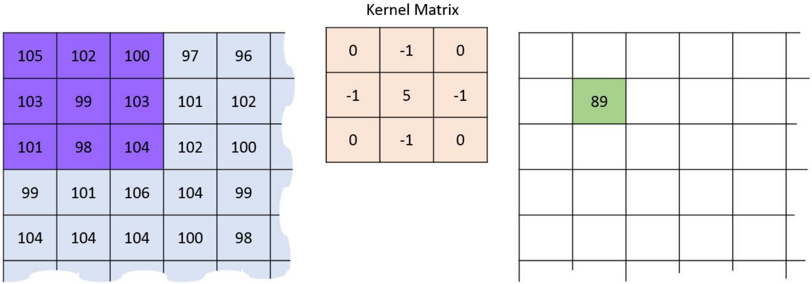


Image Matrix

$$105 * 0 + 102 * -1 + 100 * 0$$
$$+103 * -1 + 99 * 5 + 103 * -1$$
$$+101 * 0 + 98 * -1 + 104 * 0 \;\; = 89$$

Output Matrix

Filtro convolucional

Kernel Matrix

0	-1	0
-1	5	-1
0	-1	0

105	102	100	97	96
103	99	103	101	102
101	98	104	102	100
99	101	106	104	99
104	104	104	100	98

Image Matrix

$$102 * 0 + 100 * -1 + 97 * 0$$
$$+99 * -1 + 103 * 5 + 101 * -1$$
$$+98 * 0 + 104 * -1 + 102 * 0 = 111$$

	89	111		

Output Matrix

Filtro convolucional

Kernel Matrix

0	-1	0
-1	5	-1
0	-1	0

0	0	0	0	0	0
0	105	102	100	97	96
0	103	99	103	101	102
0	101	98	104	102	100
0	99	101	106	104	99
0	104	104	104	100	98

Image Matrix

$$
\begin{aligned}
&0 * 0 + 105 * -1 + 102 * 0 \\
&+0 * -1 + 103 * 5 + 99 * -1 \\
&+0 * 0 + 101 * -1 + 98 * 0 \quad = 210
\end{aligned}
$$

210	89	111		

Output Matrix

0	0	0	0	0	0
0	105	102	100	97	96
0	103	99	103	101	102
0	101	98	104	102	100
0	99	101	106	104	99
0	104	104	104	100	98

Image Matrix

Kernel Matrix

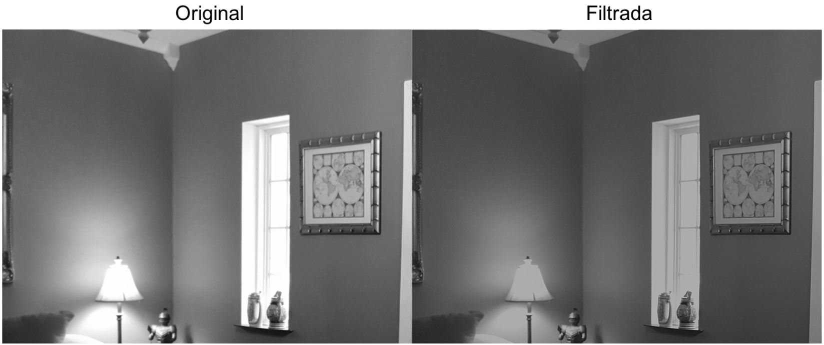
0	-1	0
-1	5	-1
0	-1	0

320				
210	89	111		

Output Matrix

$$\begin{aligned}&0*0+0*-1+0*0\\&+0*-1+105*5+102*-1\\&+0*0+103*-1+99*0 \quad =320\end{aligned}$$

Resultado



Los filtros pueden ser extremos

$$Kernel = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Y el procesamiento de imágenes de hace así. . .

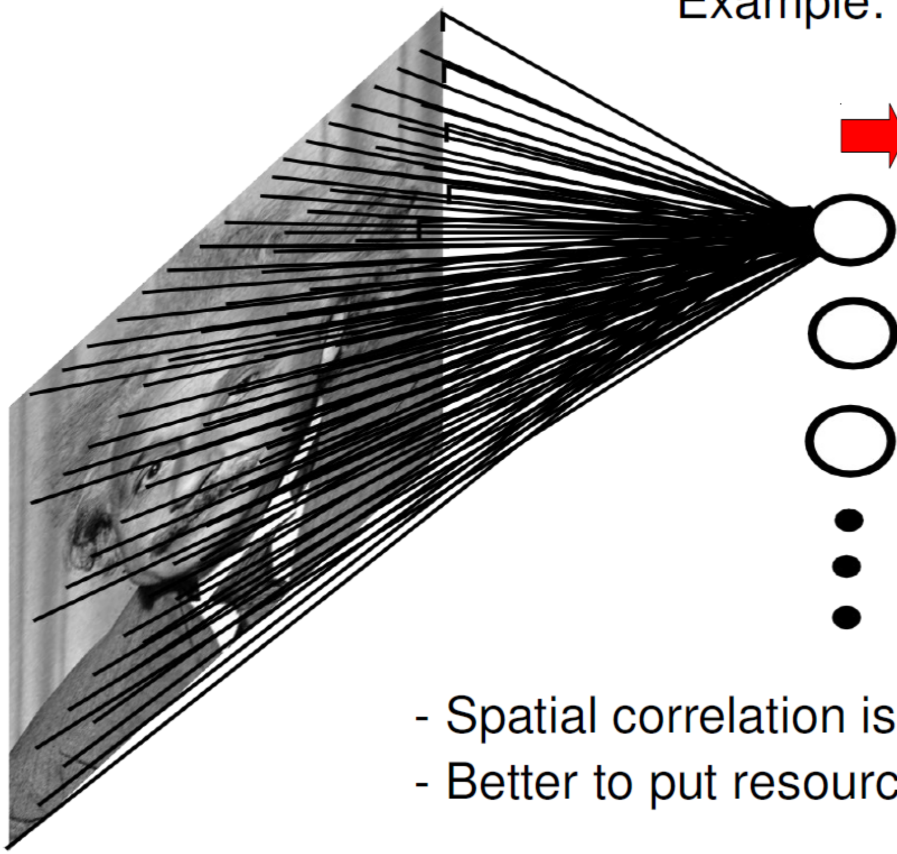
- Primero se seleccionan muchos filtros, dependiendo de lo que queremos
- Se ajustan los filtros para representar las características principales
- Se pueden aplicar filtros a las imágenes filtradas
- Se utilizan otras técnicas (submuestreo, histograma, . . .)
- Se convierte la imagen en un enorme vector de características

¿Y porquñé nos interesaría el reconocimiento de imágenes?

- Una imagen es información organizada en varias dimensiones
- Las información temporal se puede organizar en dimensiones (días/años, etc.)
- La información adjacete a un dato se asume más importante que la información más lejana
- Muchos problemas se pueden representar de esta forma (¿Cómo podrían representarse como imágenes los datos de demanda de energía?)

CNNs – key ideas

FULLY CONNECTED NEURAL NET

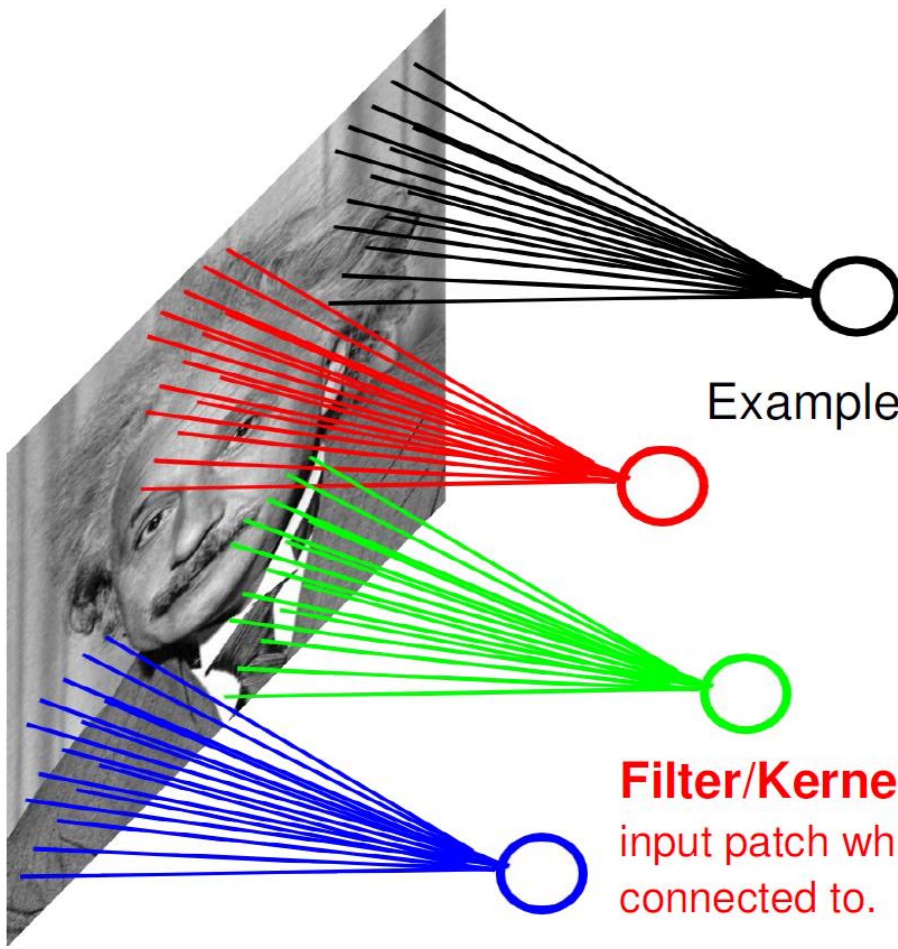


Example: 1000x1000 image
1M hidden units
10^12 parameters!!!

- Spatial correlation is local
- Better to put resources elsewhere!

CNNs – key ideas

LOCALLY CONNECTED NEURAL NET

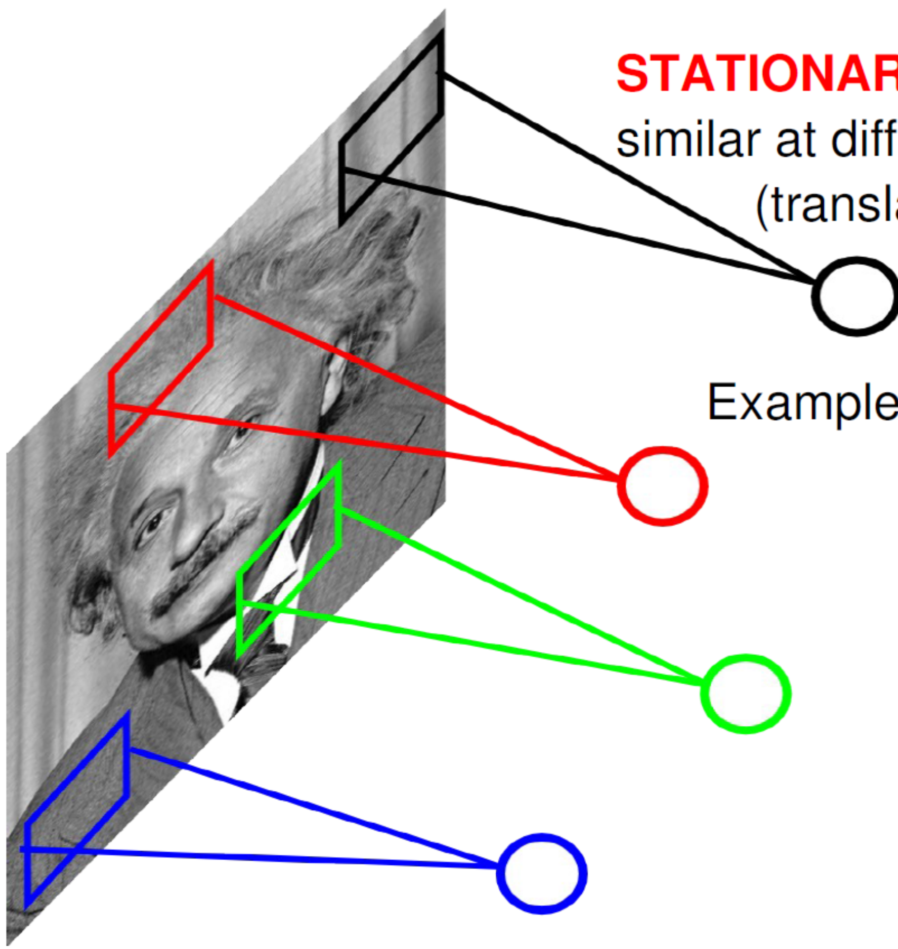


Example: 1000x1000 image
1M hidden units
Filter size: 10x10
100M parameters

Filter/Kernel/Receptive field:
input patch which the hidden unit is connected to.

CNNs – key ideas

LOCALLY CONNECTED NEURAL NET

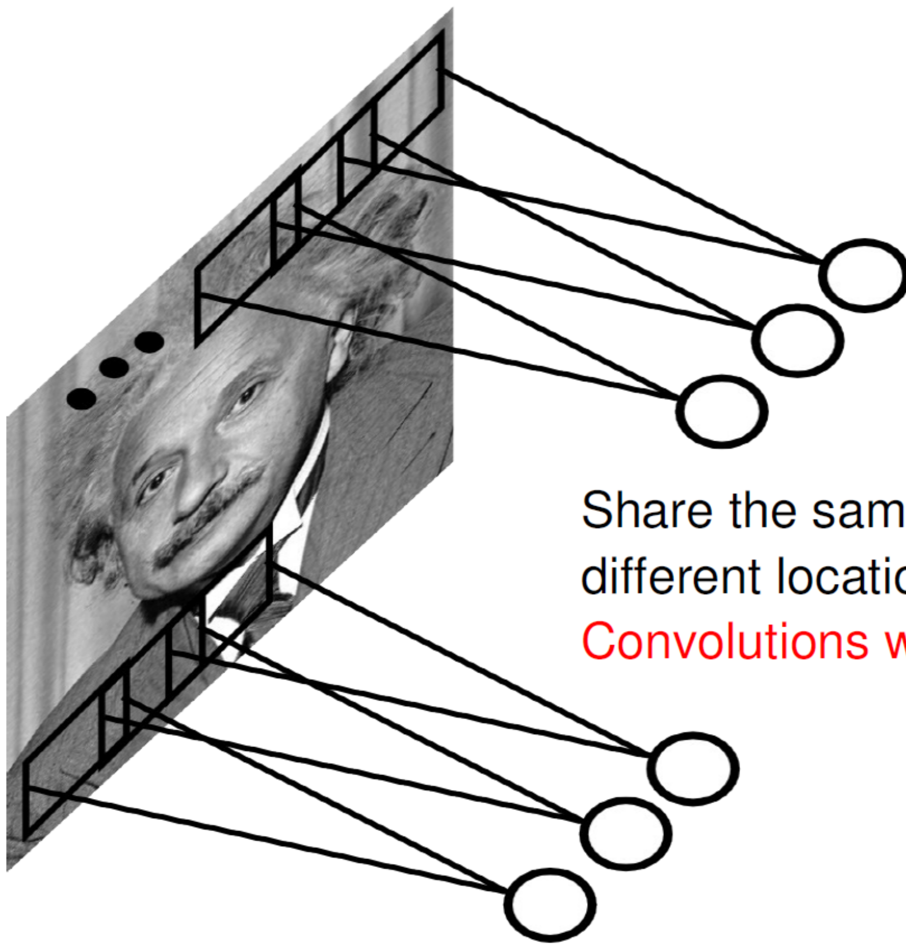


STATIONARITY? Statistics are similar at different locations (translation invariance)

Example: 1000x1000 image
1M hidden units
Filter size: 10x10
100M parameters

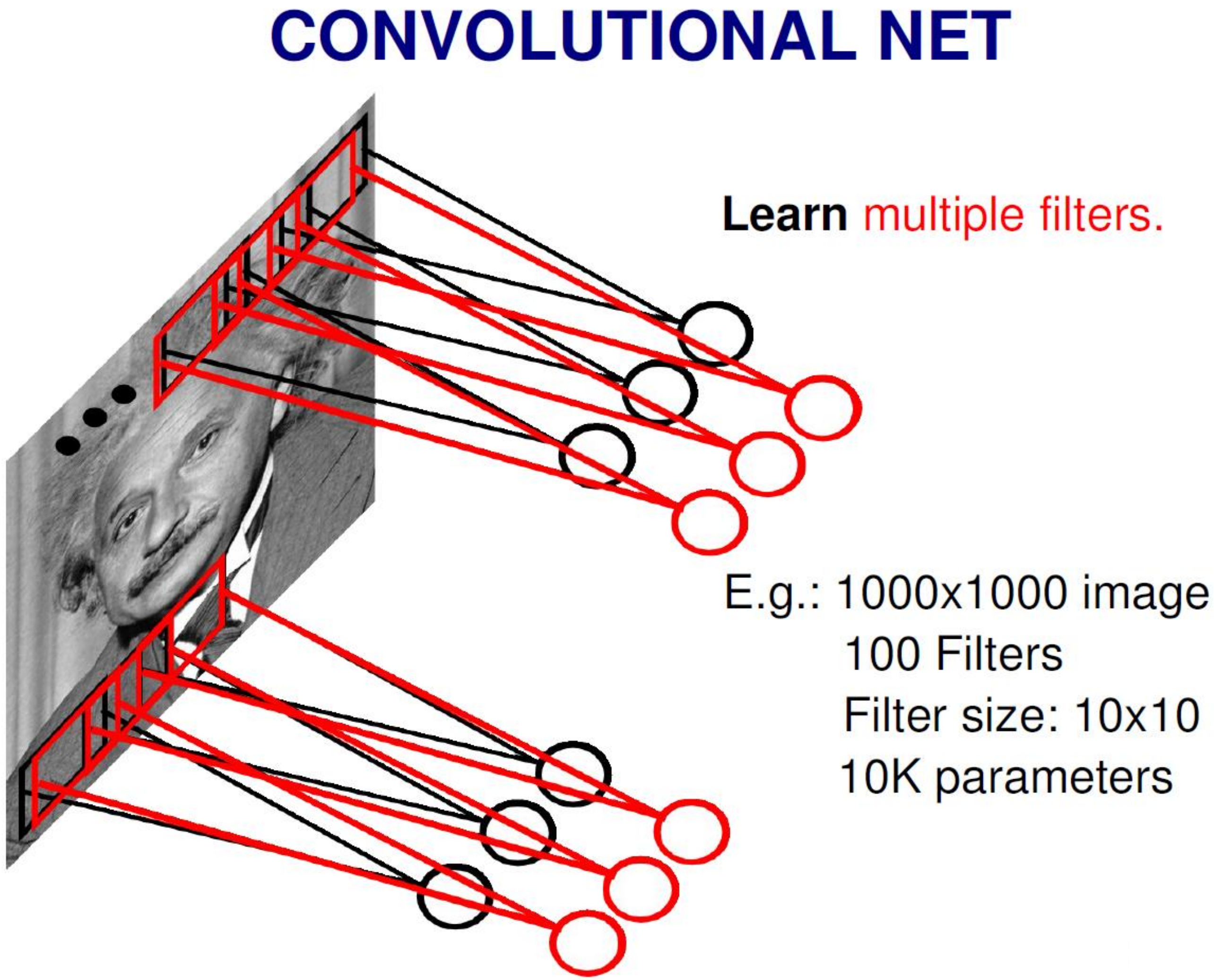
CNNs – key ideas

CONVOLUTIONAL NET

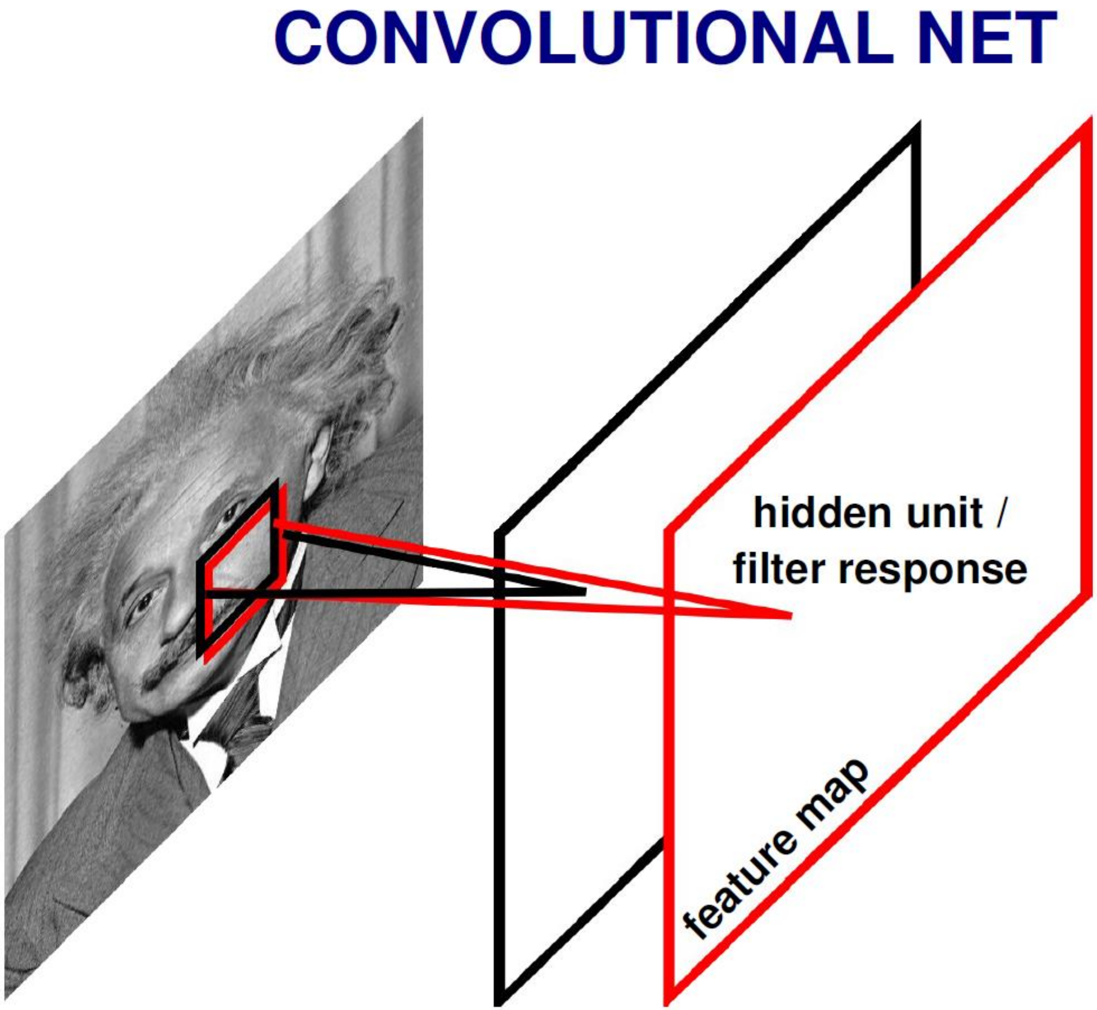


Share the same parameters across different locations:
Convolutions with learned kernels

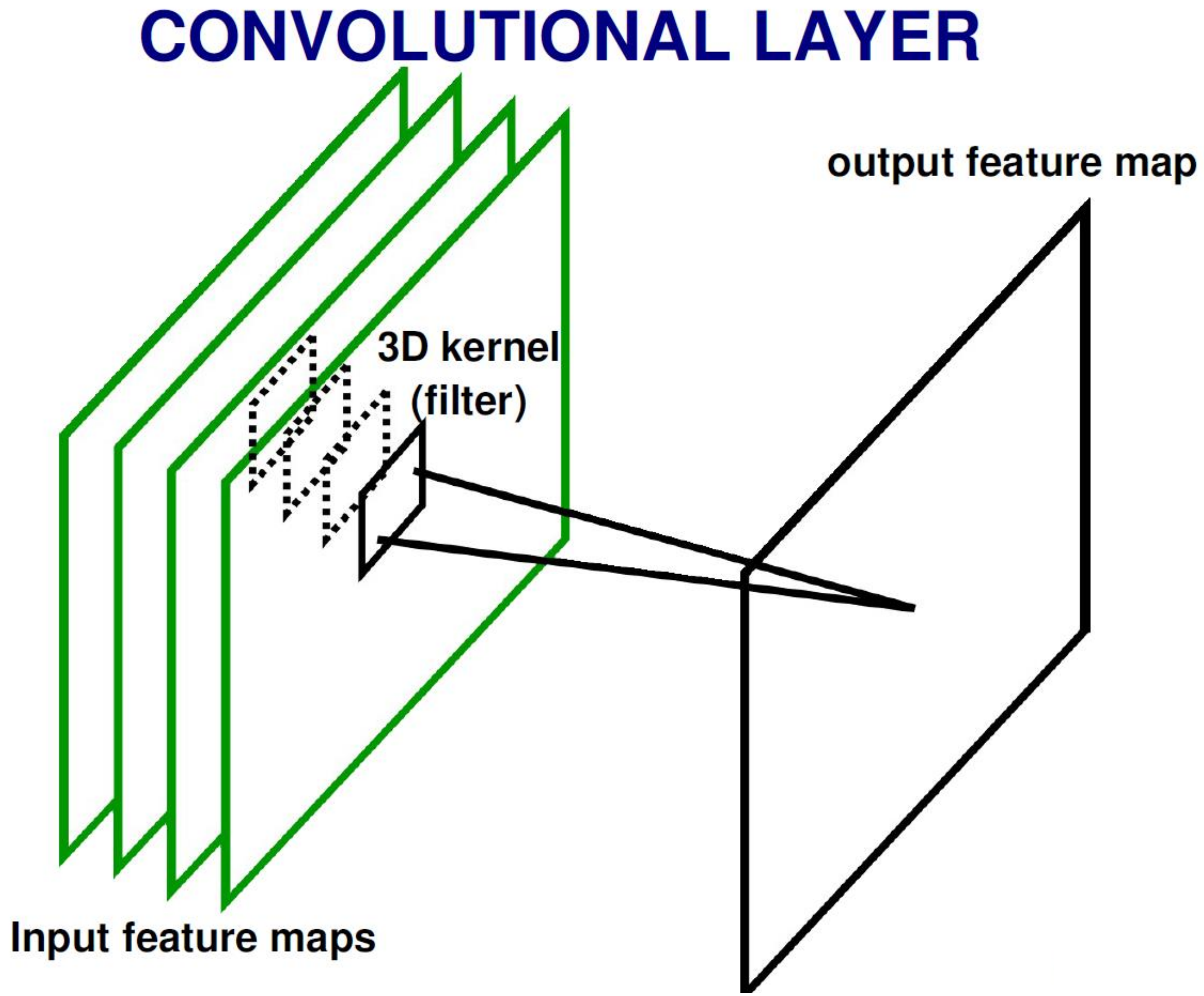
CNNs – key ideas



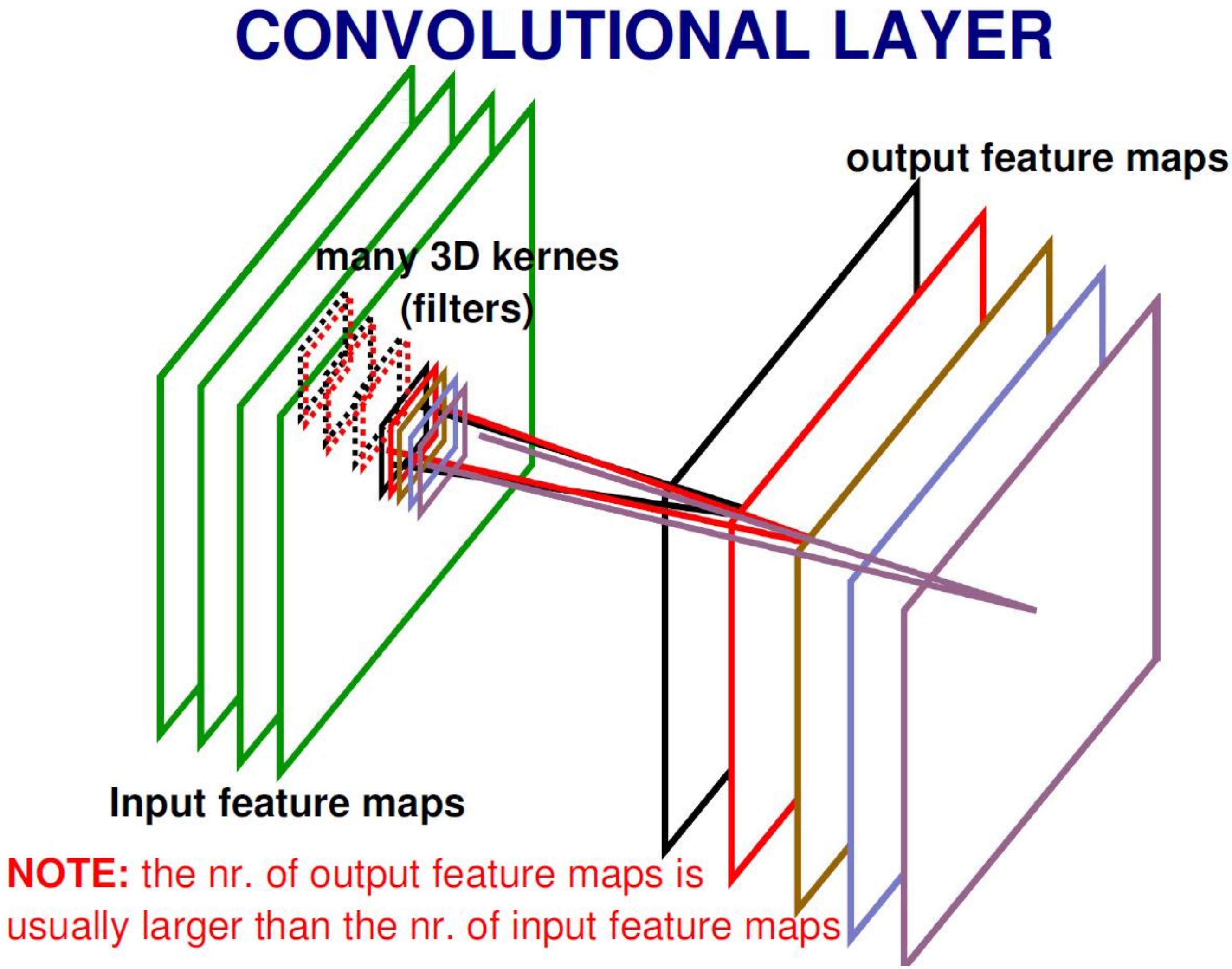
CNNs – key ideas



CNNs – key ideas

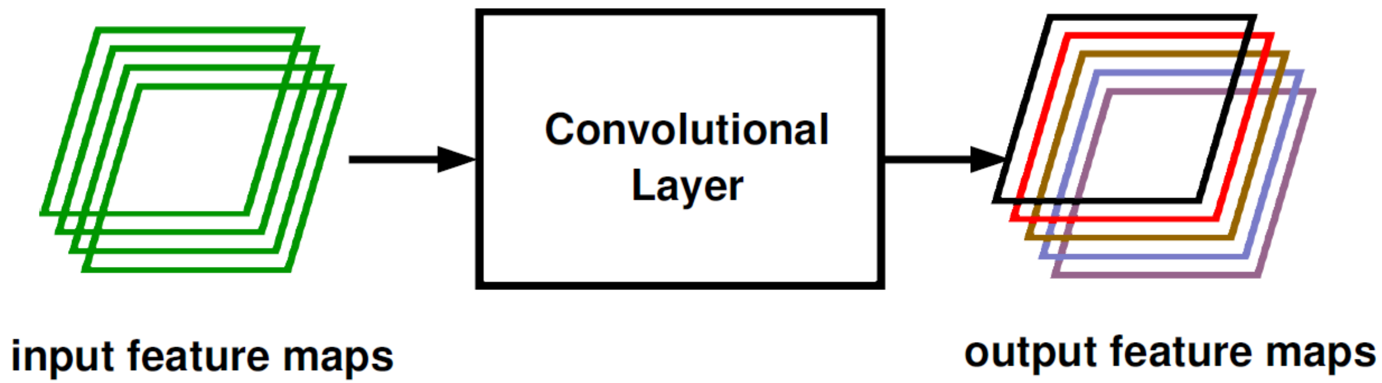


CNNs – key ideas



CNNs – key ideas

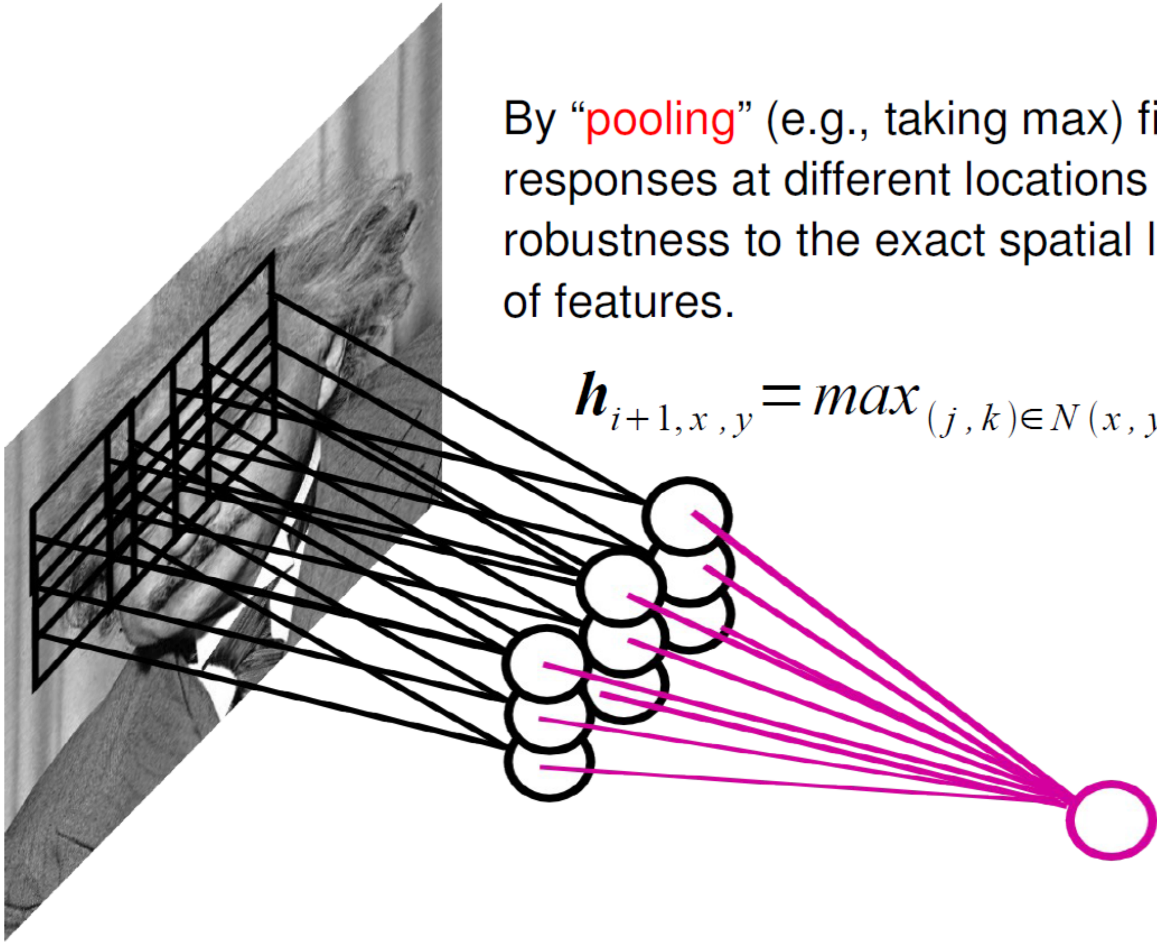
CONVOLUTIONAL LAYER



NOTE: the nr. of output feature maps is usually larger than the nr. of input feature maps

CNNs – key ideas

POOLING

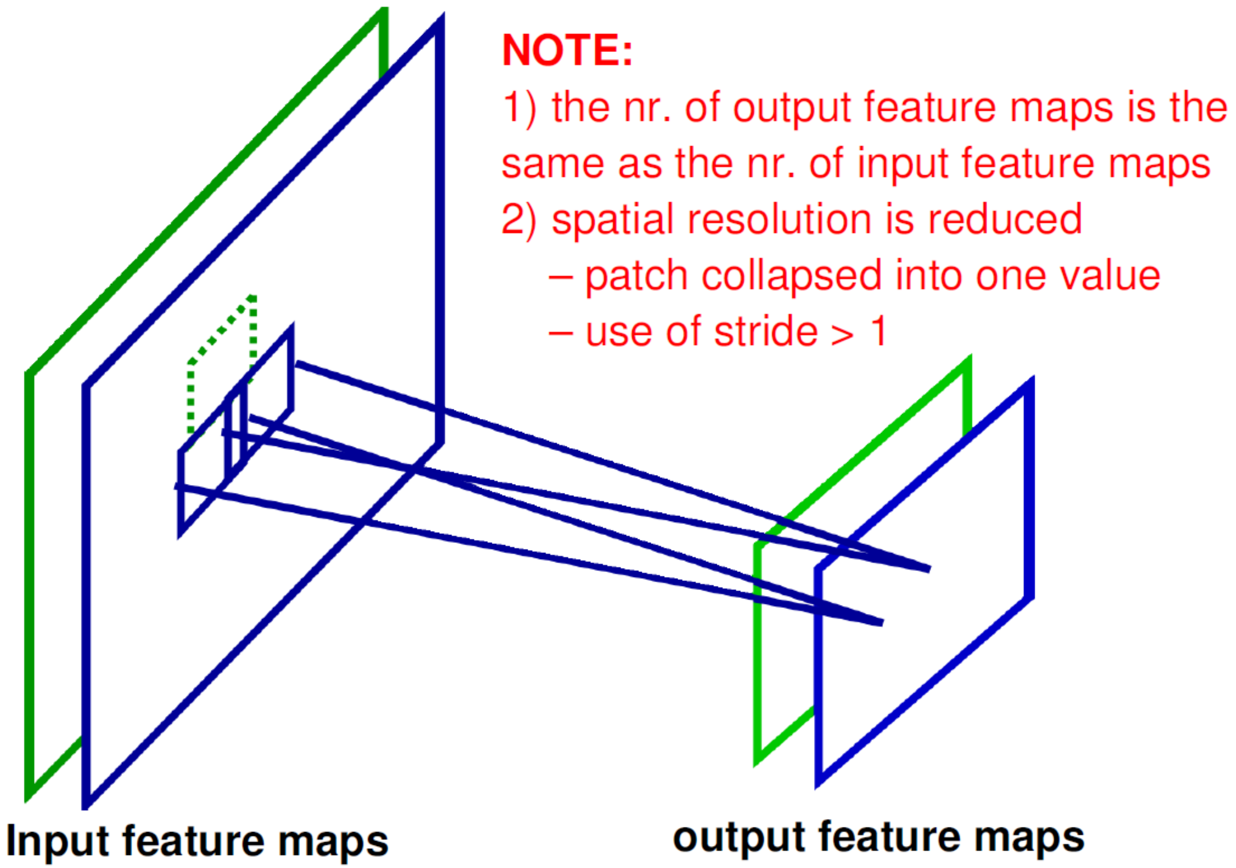


By "pooling" (e.g., taking max) filter responses at different locations we gain robustness to the exact spatial location of features.

$$h_{i+1,x,y} = max_{(j,k)\in N(x,y)} h_{i,j,k}$$

CNNs – key ideas

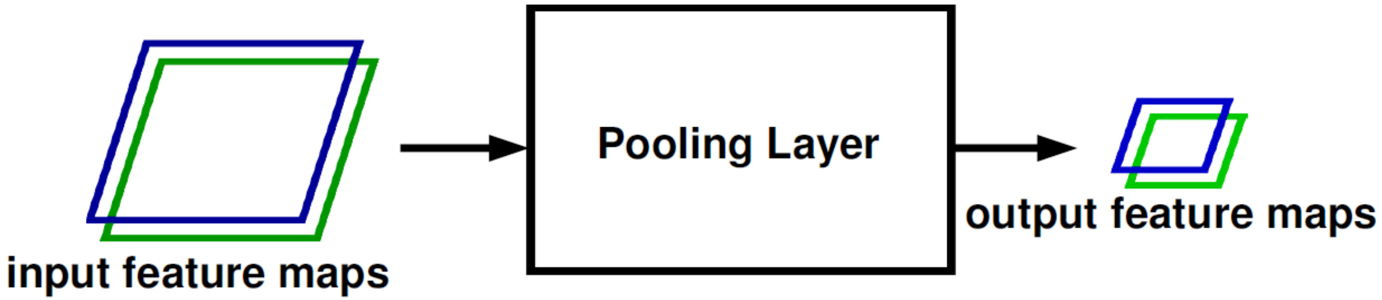
POOLING LAYER



CNNs – key ideas

POOLING LAYER

NOTE:
1) the nr. of output feature maps is the same as the nr. of input feature maps
2) spatial resolution is reduced
   – patch collapsed into one value
   – use of stride > 1



CNNs – typical architecture

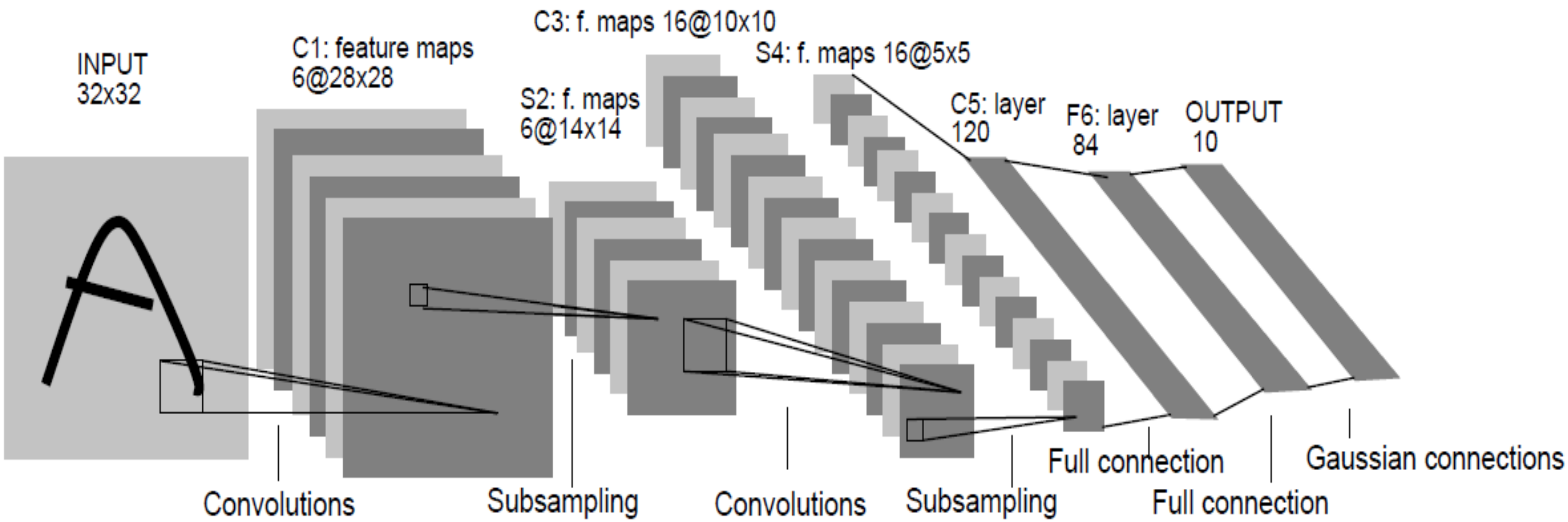
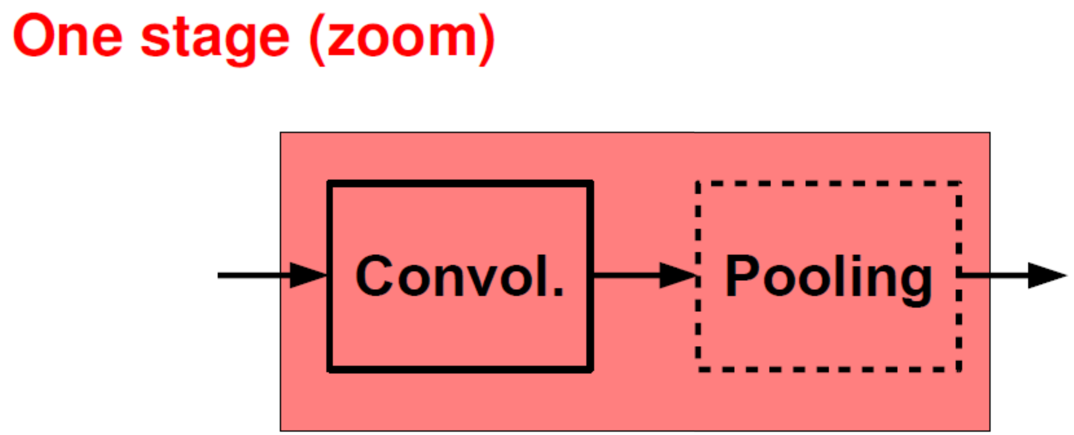
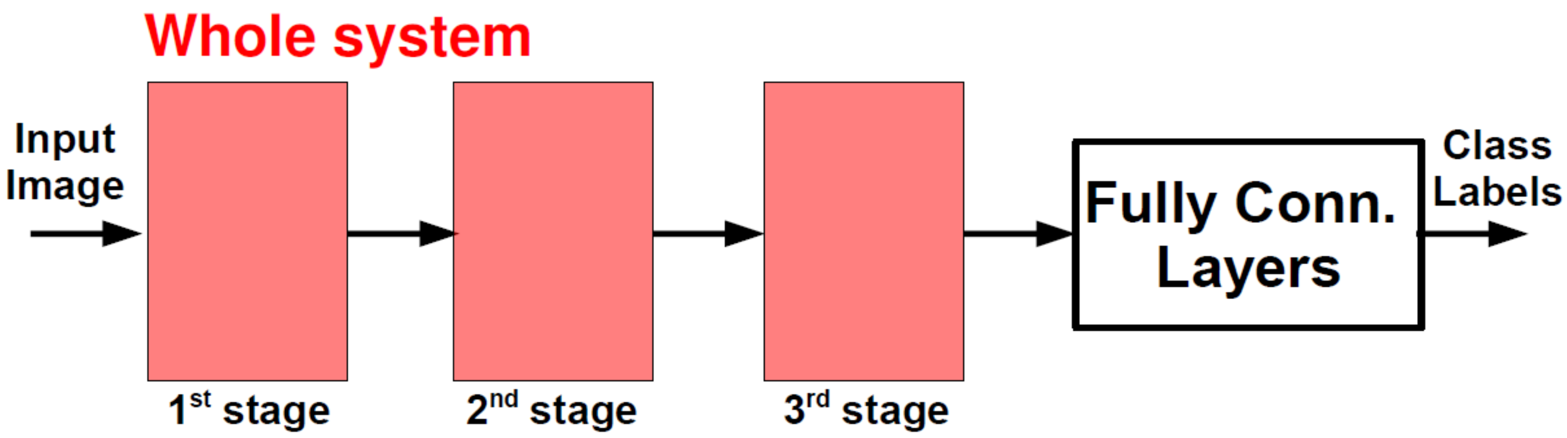


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

CNNs – typical architecture

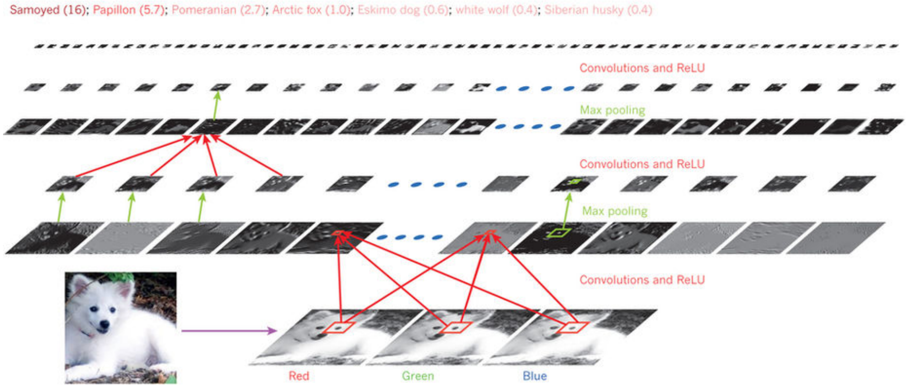




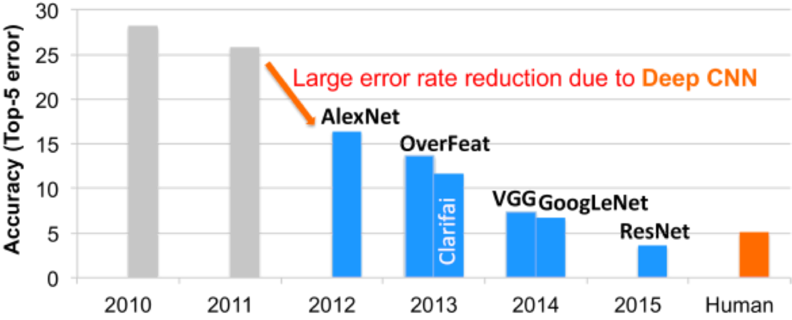
CNNs – conclusion

- Connect each hidden unit to a small patch of the input.
- Share the weight across hidden units.
- Subsampling layers are useful to reduce computational burden and increase invariance.

Redes convolucionales (CNN)
Arquitectura general



Resultados aplicados al conjunto de datos *ImageNet*



Muchas gracias por su atención