

Indice

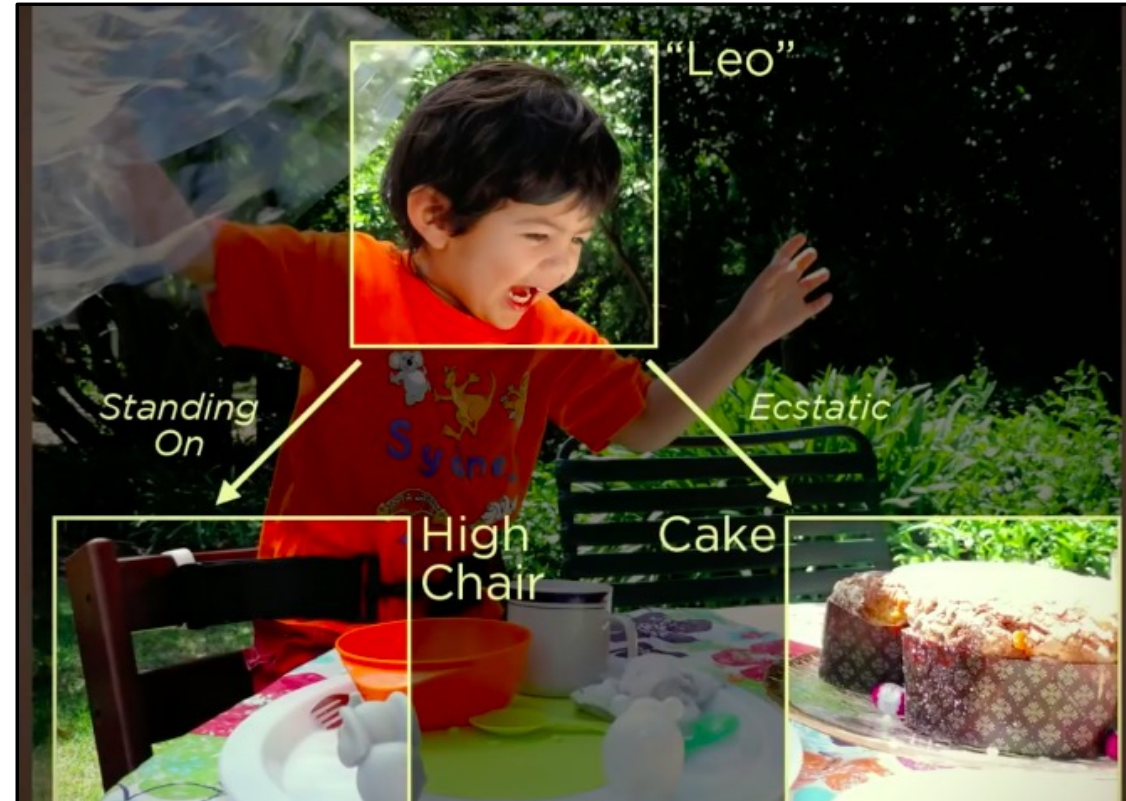
- Introduzione
- Convoluzione nella computer vision
- Padding, Stride
- Convoluzione su volumi



CNN – Convolutional Neural Network

- Una CNN è una **rete neurale** con almeno un livello convolutivo
- Nate ispirandosi dalla organizzazione della corteccia visiva animale
- Utilizzate in:
 - *Riconoscimento in immagini e video*
 - *Sistemi di raccomandazione*
 - *NLP*
 - *Bioinformatica*
 - *Analisi segnali*

Principale peculiarità: **pattern recognition**

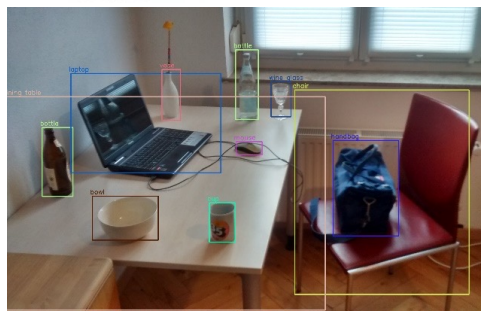


CNN – Utilizzo tipico

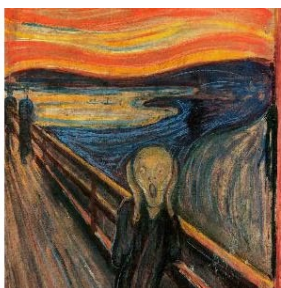


64x64

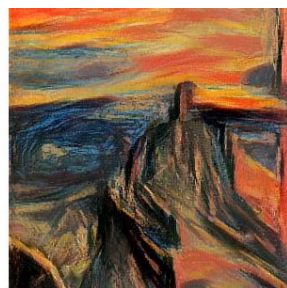
→ Cat? (0/1)



+



=



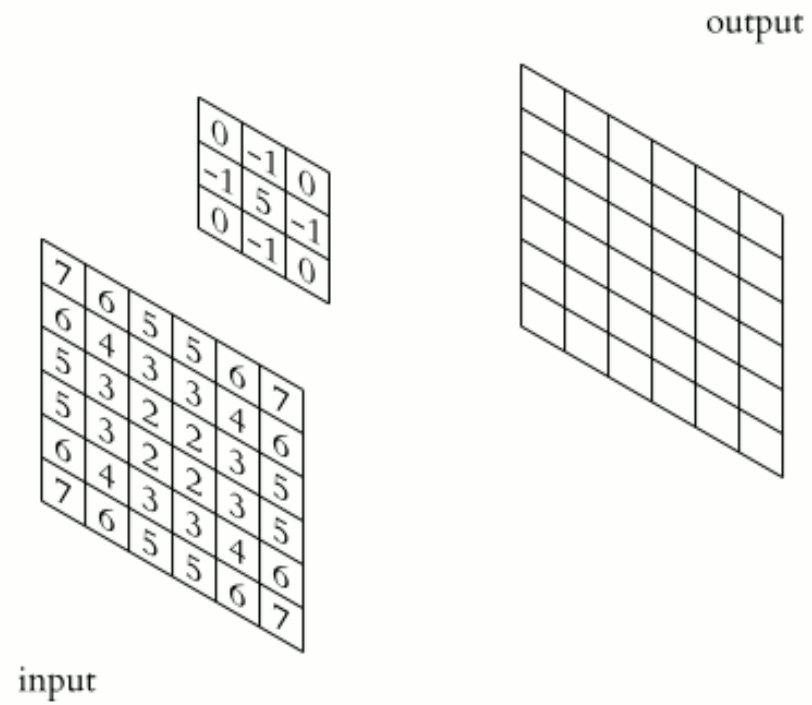
- Classificazione di immagini
- Riconoscimento automatico di oggetti
- Trasferimento di stili
- Posizione di oggetti in un'immagine
- Segmentazione di un'immagine
- Riconoscimento pattern ricorrenti in segnali
- Classificazione testi
- Analisi serie temporali di dati

Convoluzione

- È una tecnica molto usata nella computer vision per effettuare dei **filtraggi** (es. Blur)
- Prende un'immagine in input e restituisce un'immagine in output
- Prevede l'utilizzo di un *kernel* o *filtro* di dimensione contenuta (es. 3x3, 5x5)
- Il kernel viene sovrapposto ad ogni gruppo di pari dimensioni dell'immagine originaria
- Vengono effettuate delle operazioni matematiche (prodotti e somme algebriche)
- I risultati sono i pixel della nuova immagine
- A seconda del kernel utilizzato è possibile estrarre *feature* diverse



Convoluzione



Convoluzione

<https://setosa.io/ev/image-kernels/>



Padding

Ipotesi

Ogni pixel di un'immagine di dimensioni $W \times H$ viene convoluta con un kernel di dimensioni $W_K \times H_K$:

Risultato

L'immagine finale ha dimensioni $(W - W_K + 1) \times (H - H_K + 1)$

- Es.
 - Un'immagine ha dim. 6×6
 - Dim. Filtro: 3×3
 - l'immagine in output avrà dimensioni 4×4
- Questo accade perché i pixel della cornice non danno contributi in una convoluzione

Padding

120	86	55	78	33	12
12	132	24	255	129	122
255	255	24	255	128	128
255	44	11	24	128	224
35	25	2	20	22	141
52	51	28	5	52	14

Input

-1	0	1
-1	0	1
-1	0	1

Kernel

-284	115	187	-326
-463	...		

Output

Padding

- Per convoluzioni successive si potrebbe arrivare a dimensioni negative
- Si "allarga" l'immagine di una certa quantità di pixel settati a 0, detto **padding**

• Es. $p = 1$

0	0	0	0	0	0	0	0
0	120	86	55	78	33	12	0
0	12	132	24	255	129	122	0
0	255	255	24	255	128	128	0
0	255	44	11	24	128	224	0
0	35	25	2	20	22	141	0
0	52	51	28	5	52	14	0
0	0	0	0	0	0	0	0

Input

-1	0	1
-1	0	1
-1	0	1

Kernel

-218	53	-115	-83	152	...

Output



Padding

- Una convoluzione che preserva la dimensione dell'immagine in input prende il nome di **same**

$$\text{padding} = (\text{dimens. Filtro dispari} - 1) / 2$$

- Una convoluzione che fa downsampling dell'immagine in input prende il nome di **valid**

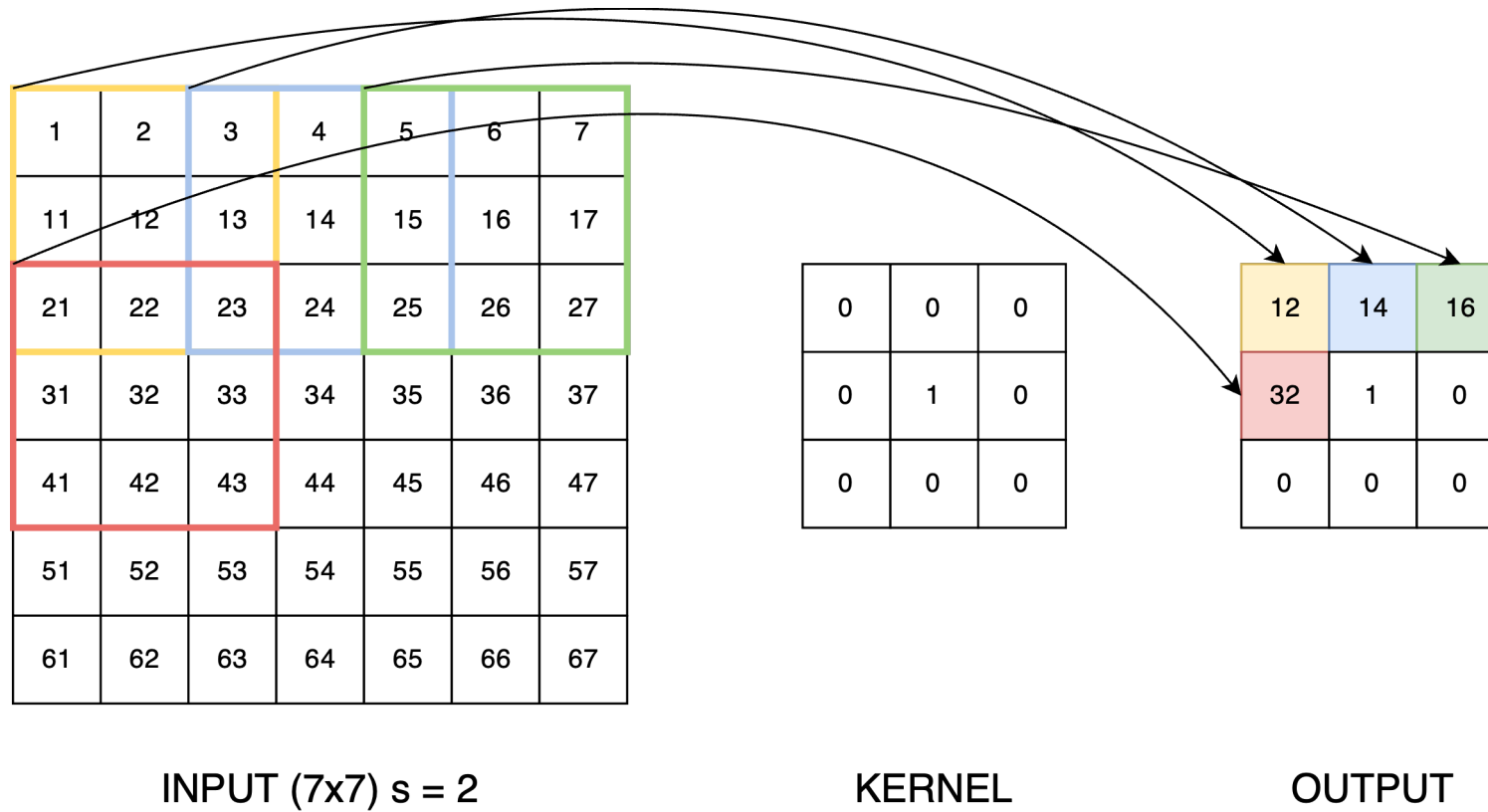
$$\text{padding} = 0$$

Stride

- Indica di quanto si sposta il kernel mentre scorre le immagini
- $s = 1$ significa che il kernel viene traslato un pixel per volta
- $s = 2$ significa che il kernel "salta" un pixel (si sposta di 2 pixel per volta)

N.B. il kernel si sposta di una quantità pari a s anche in verticale

Stride



Convoluzione volumetrica



R channel



G channel



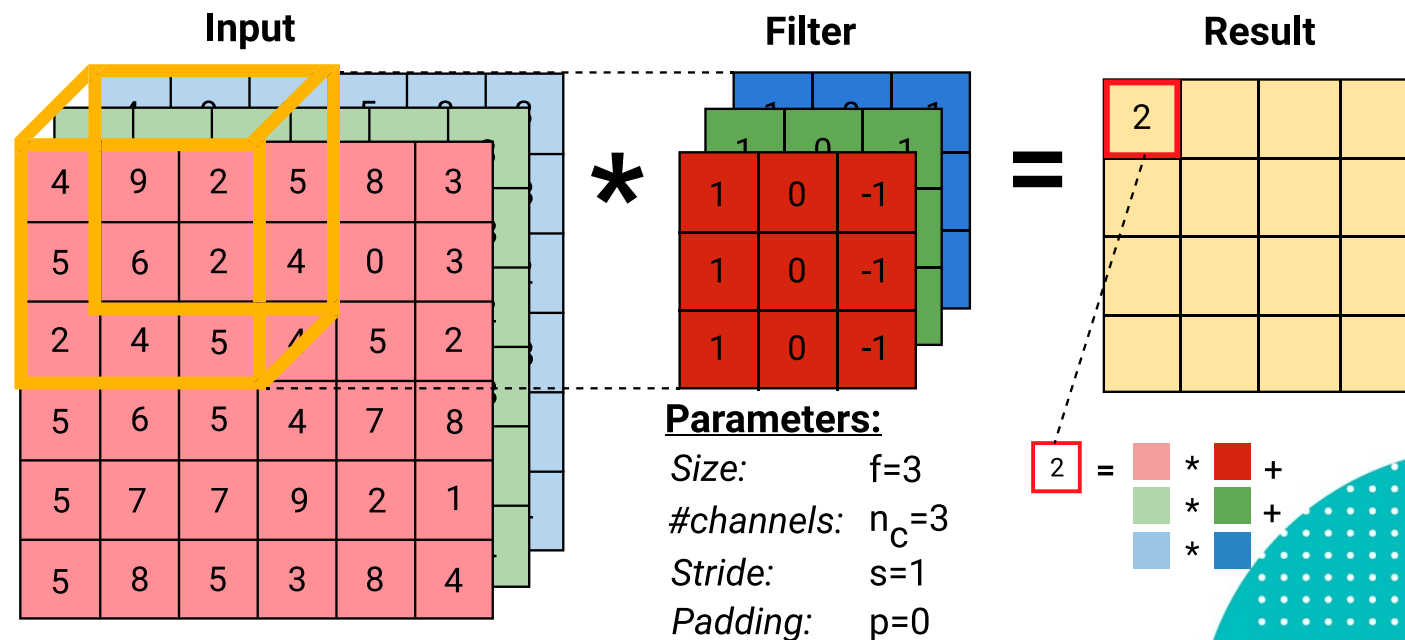
B channel

Che **succede**
se abbiamo **più canali**
(es. *Più colori?*)

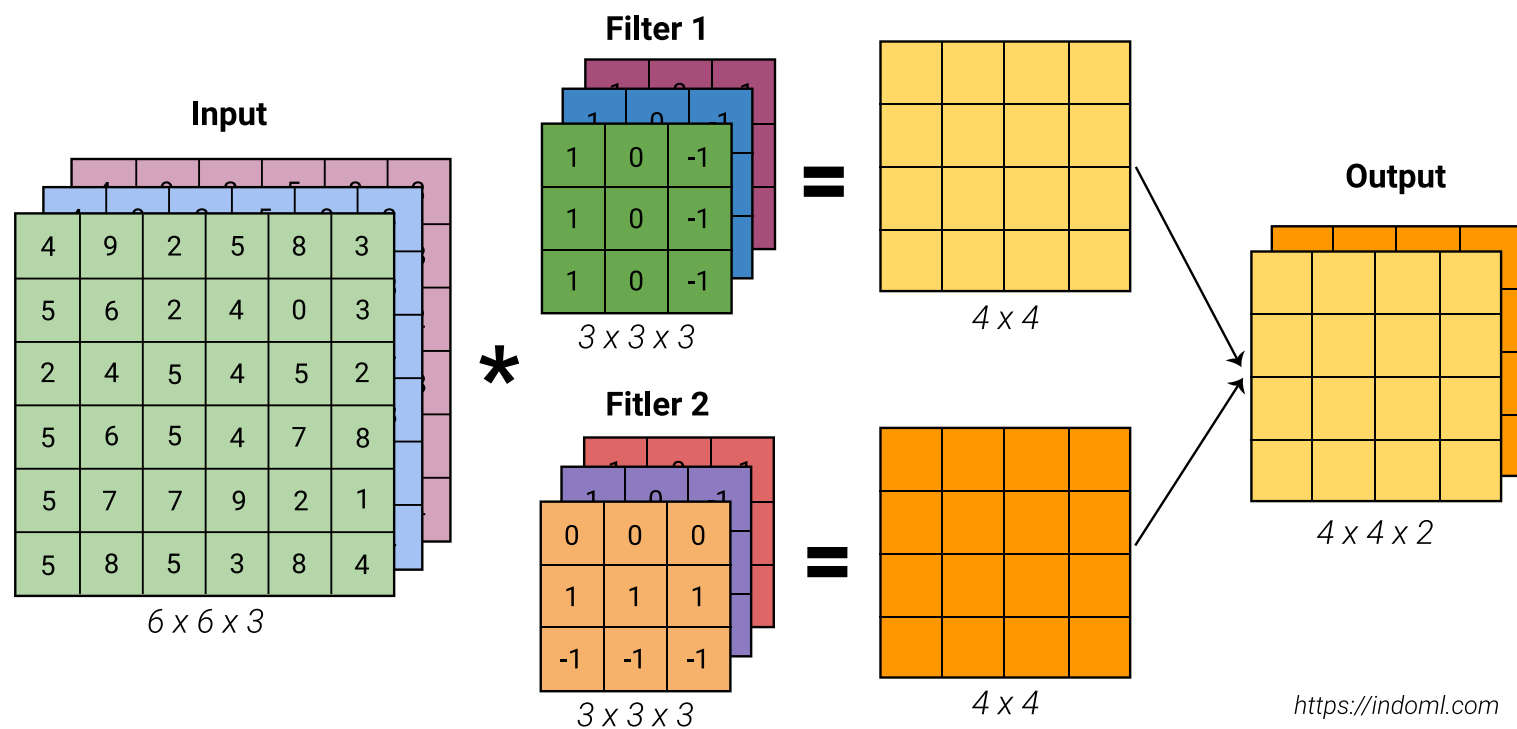
Es. Immagini RGB
(Red, Green, Blue)

Convoluzione volumetrica

- Si associa un kernel quadrato **per ogni canale** dell'immagine che abbiamo
- Un filtro è composto da un numero di kernel **uguale al numero di canali** dell'input
- Quindi abbiamo **un kernel per canale** (un kernel opera solo su un canale)



Convoluzione volumetrica



<https://indoml.com>

In caso di **numero di filtri** maggiore di 1, l'output sarà multicanale