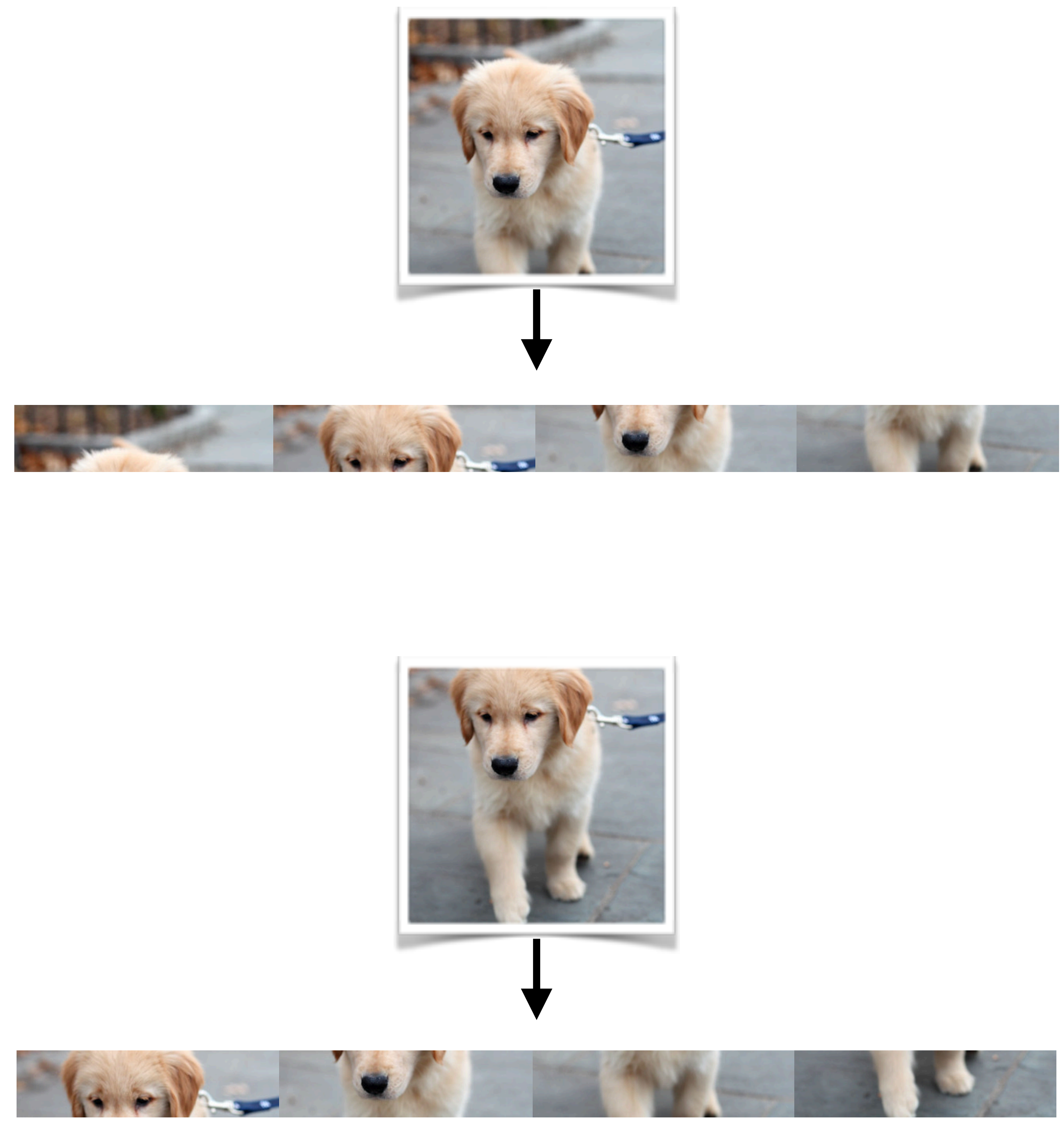
Convolutions

Images and structure

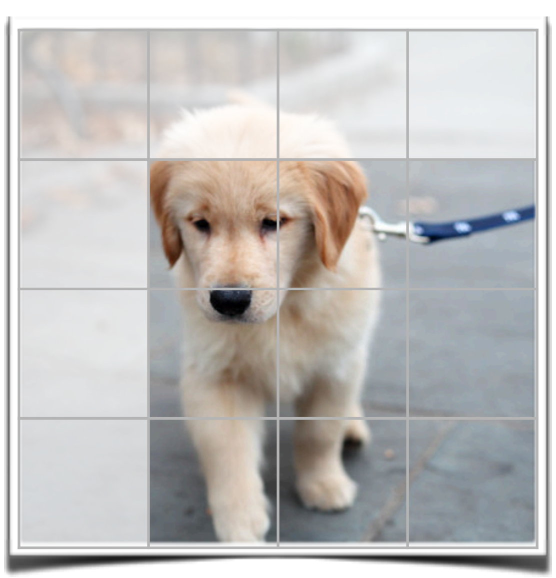
- Fully connected networks are not shift invariant



Finding shift-invariant patterns

Convolutions

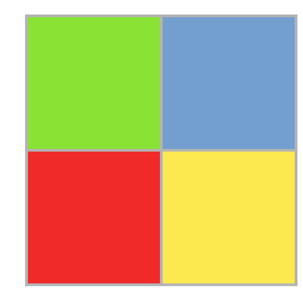
- "Sliding" linear transformation



*

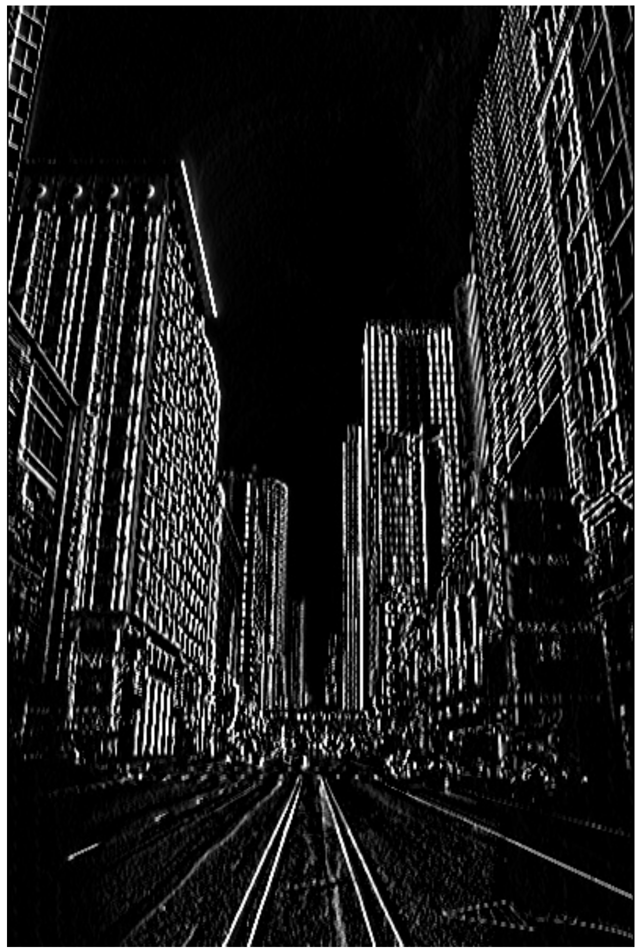
a	b	c
d	e	f
g	h	i

=



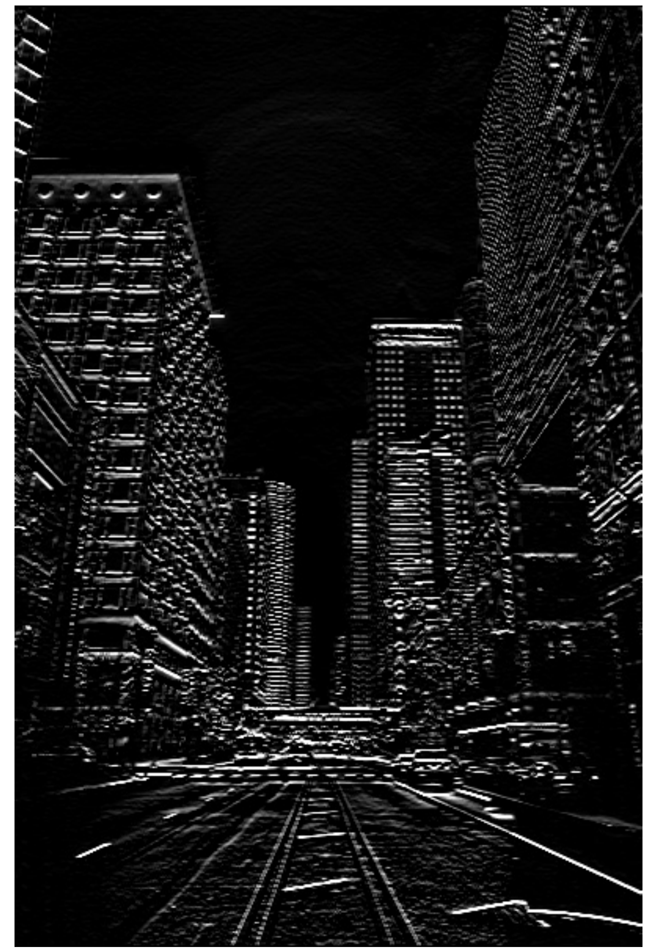
Examples of convolutions

Original



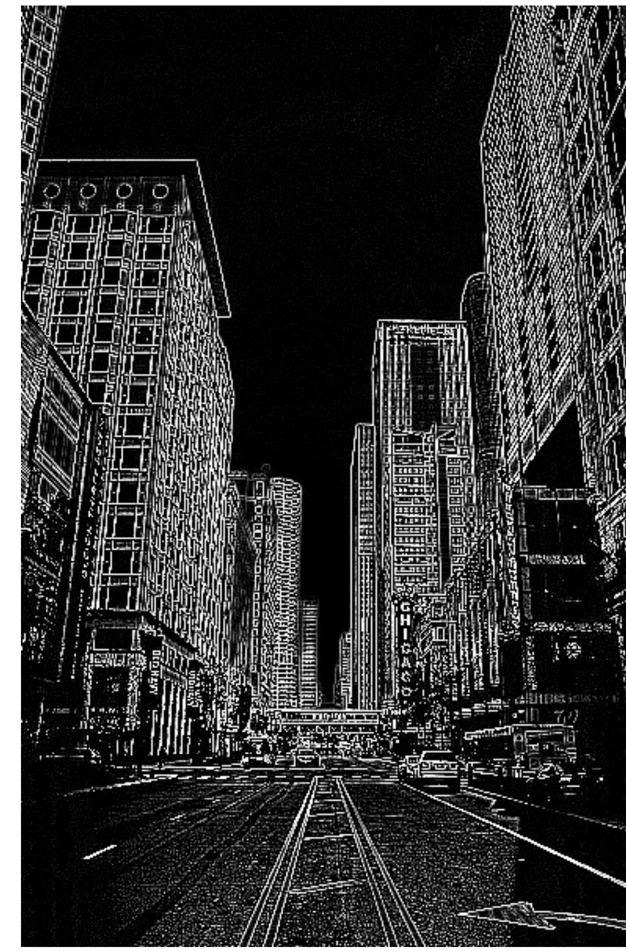
Vertical edges

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



Horizontal edges

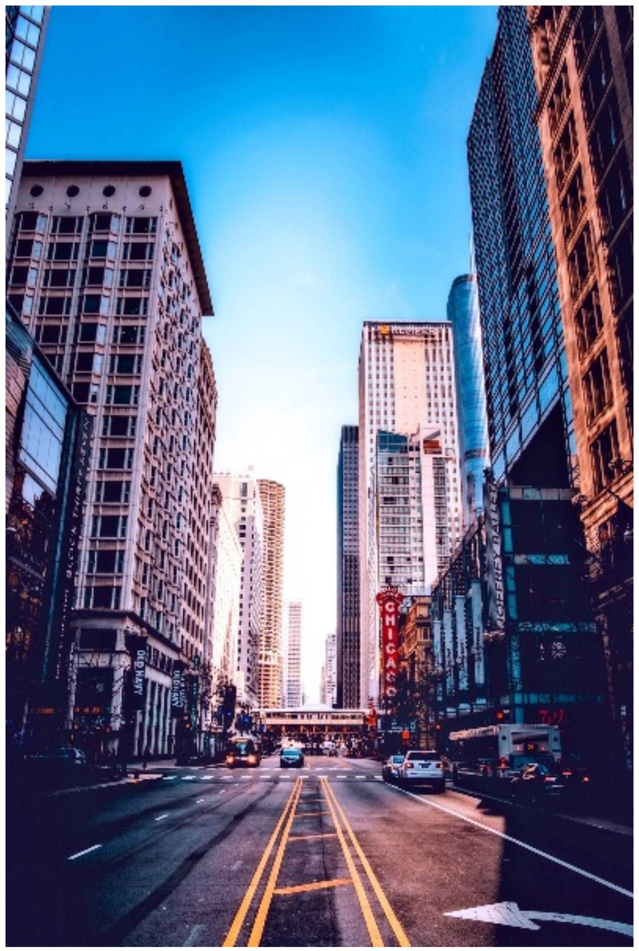
-1	-1	-1
0	0	0
1	1	1



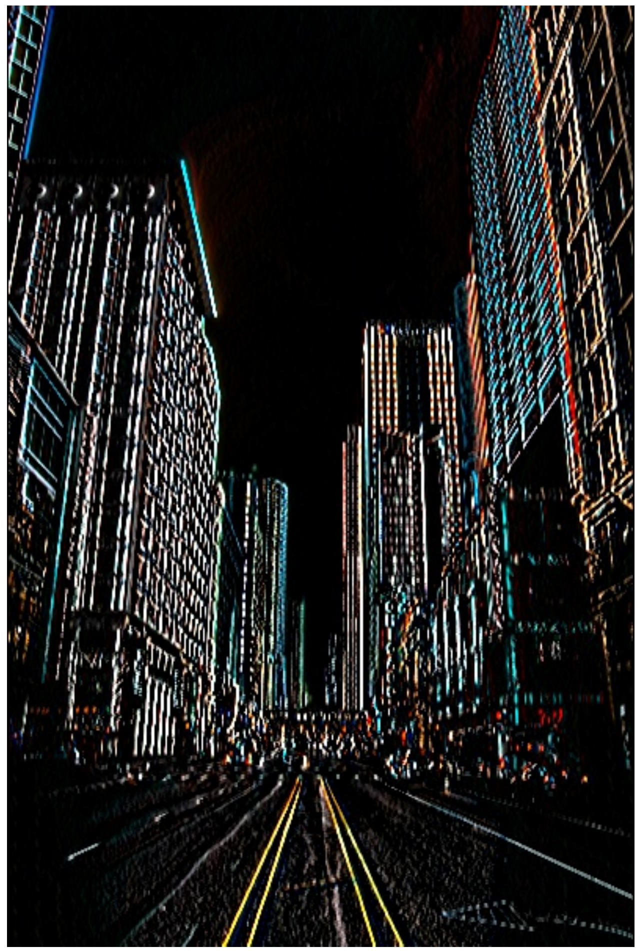
Laplace filter

-1	-1	-1
-1	8	-1
-1	-1	-1

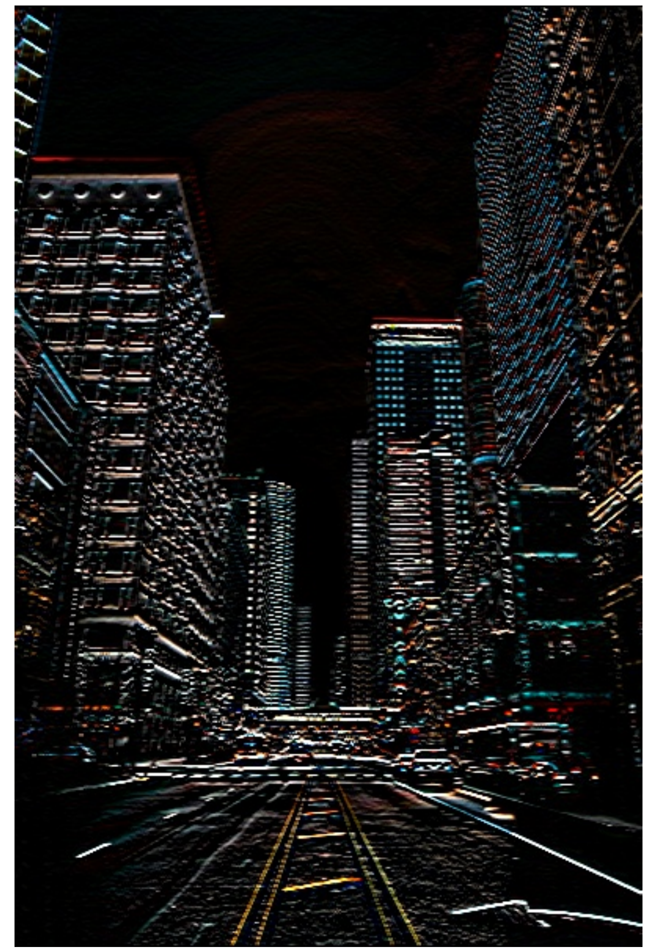
Convolutions on multiple channels



Original



Vertical edges



Horizontal edges

Laplace filter

Vertical edges:

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

Horizontal edges:

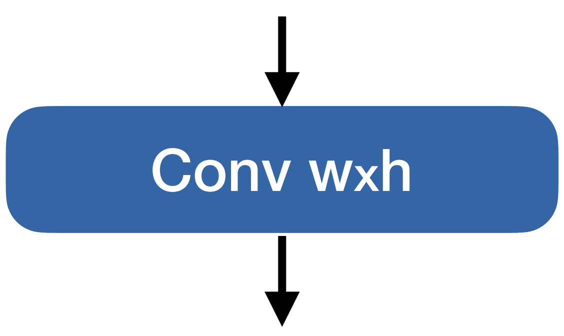
-1	-1	-1
0	0	0
1	1	1

Laplace filter:

-1	-1	-1
-1	8	-1
-1	-1	-1

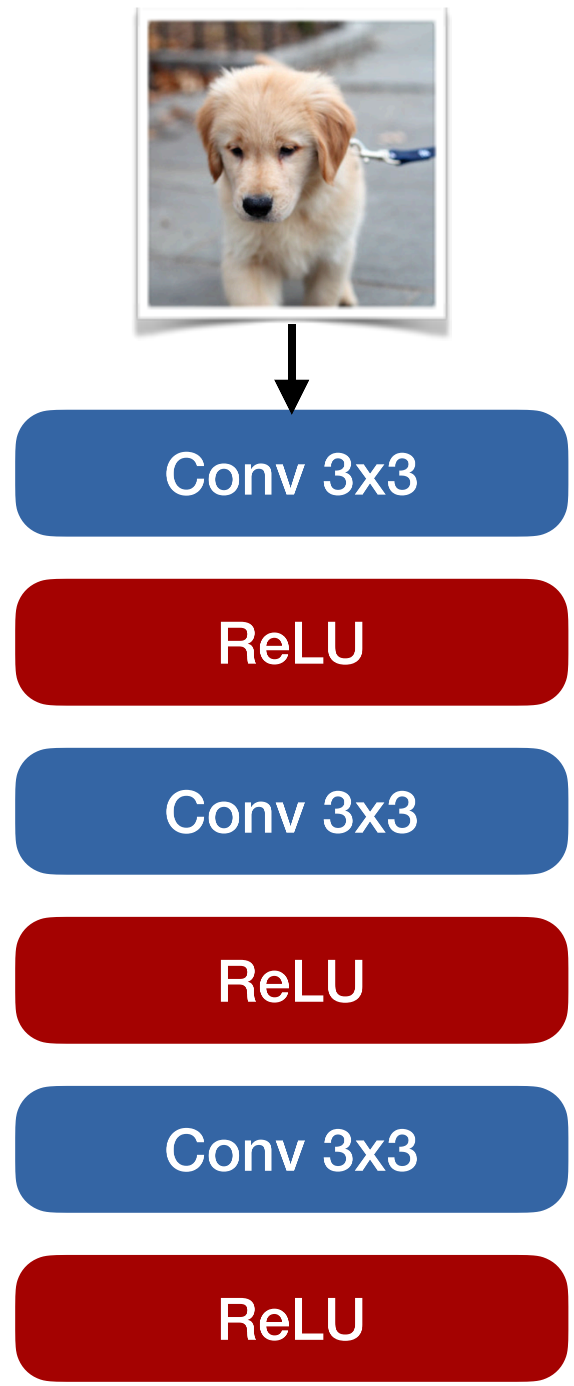
Formal definition

- Input: $\mathbf{X} \in \mathbb{R}^{H \times W \times C_1}$
- Kernel: $\mathbf{w} \in \mathbb{R}^{h \times w \times C_1 \times C_2}$
- Bias: $\mathbf{b} \in \mathbb{R}^{C_2}$
- Output: $\mathbf{Z} \in \mathbb{R}^{(H-h+1) \times (W-w+1) \times C_2}$



$$\mathbf{Z}_{a,b,c} = \mathbf{b}_c + \sum_{i=0}^{h} \sum_{j=0}^{w} \sum_{k=0}^{C_1} \mathbf{X}_{a+i,b+j,k} w_{i,j,k,c}$$

Stacking multiple layers



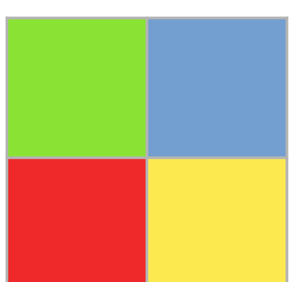
Convolution as a linear layer

input: 3 x 3 x 1

kernel: 2 x 2 x 1 x 1

a	b
c	d

output: 2 x 2 x 1

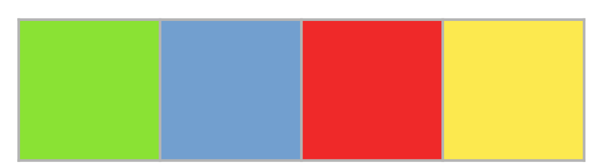


input: 9

weight: 4 x 9

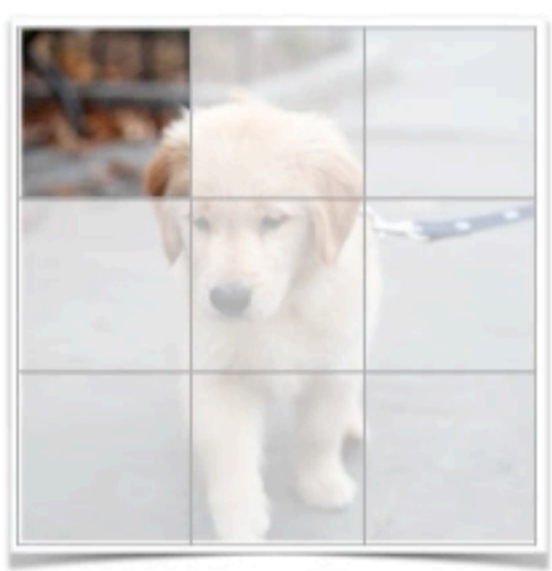
a	b		c	d				
	a	b		c	d			
			a	b		c	d	
				a	b		c	d

output: 4



Special case: 1x1 convolution

- Pixel-wise linear transformation
- Kernel: $1 \times 1 \times C_1 \times C_2$



$*$ $\mathbf{W}$ =