Convolutional neural networks

Bálint Ármin Pataki

L-layer neural network: reminder

$x \in \mathbb{R}^N, y \in \mathbb{R}^K$, neural network: $\mathbb{R}^N \to \mathbb{R}^K$

$$z^{[1]} = W^{[1]}x + b^{[1]}, \quad W: n^{[1]} \times N, \quad b: n^{[1]} \times 1$$
$$a^{[1]} = g\left(z^{[1]}\right)$$

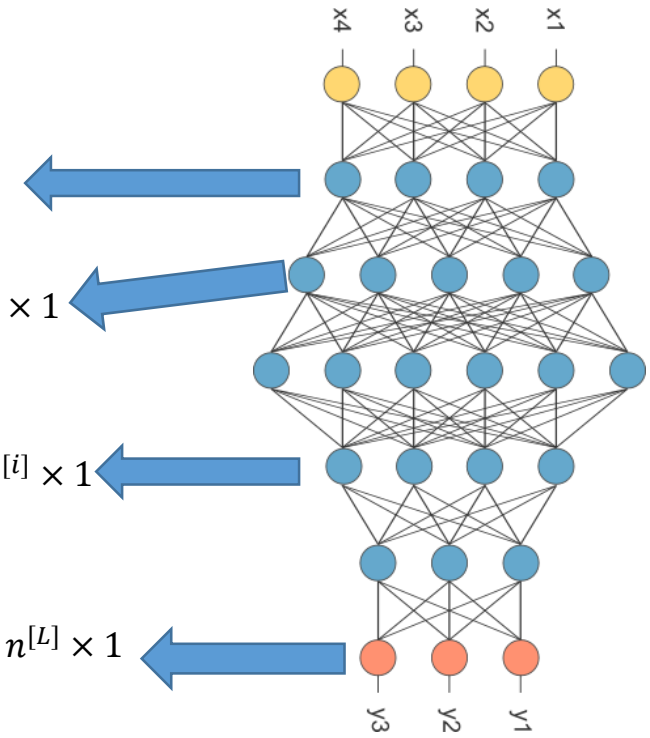
$$z^{[2]} = W^{[2]}a^{[1]} + b^{[2]}, \quad W: n^{[2]} \times n^{[1]}, \quad b: n^{[2]} \times 1$$
$$a^{[2]} = g\left(z^{[2]}\right)$$

$$\vdots$$

$$z^{[i]} = W^{[i]}a^{[i-1]} + b^{[i]}, \quad W: n^{[i]} \times n^{[i-1]}, \quad b: n^{[i]} \times 1$$
$$a^{[i]} = g\left(z^{[i]}\right)$$

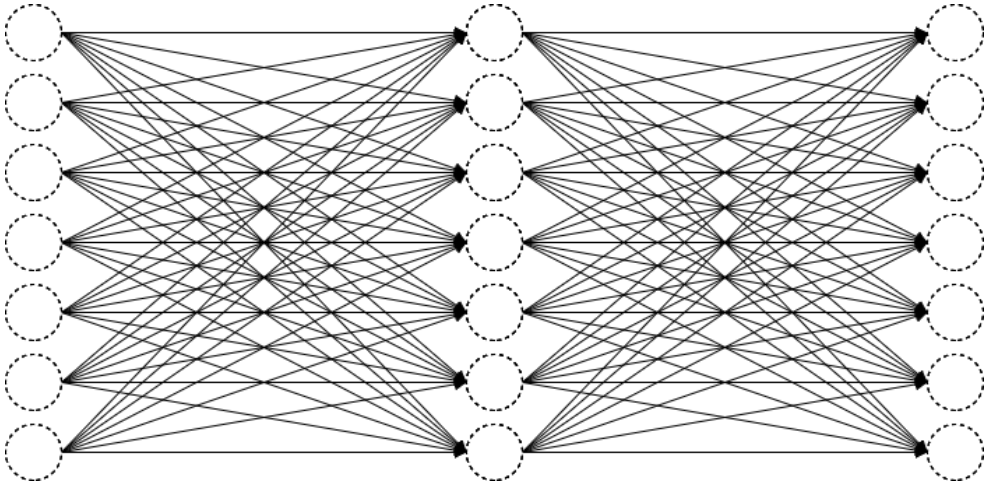
$$\vdots$$

$$z^{[L]} = W^{[L]}a^{[L-1]} + b^{[L]}, \quad W: n^{[L]} \times n^{[L-1]}, \quad b: n^{[L]} \times 1$$
$$y = a^{[L]} = softmax\left(z^{[L]}\right)$$



Credit: OpenNN

Dense neural networks: problems for real world images 1.



- Exploding parameter number:
  - 200x200 pixel input → 40000 input
  - $40000^2 + 40000 \approx 1.6 \cdot 10^9$ parameters per layer
  - float32: 4 byte/number → 6.4 GB/layer
  - color images have 3 color channels (RGB) → 57.6 GB/layer

Dense neural networks: problems for real world images 2.

Translation invariance
- same object appears at different part of the image is still the same object
- shared information -- 'tree detector', 'dog detector'

https://cs.stanford.edu/people/karpathy/ilsvrc/

Convolution in deep learning

filter

w_{00}	w_{01}	w_{02}
w_{10}	w_{11}	w_{12}
w_{20}	w_{21}	w_{22}

Image

a_{00}	a_{01}	a_{02}	a_{03}	a_{04}	a_{05}
a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
a_{20}	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}
a_{30}	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}
a_{40}	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}
a_{50}	a_{51}	a_{52}	a_{53}	a_{54}	a_{55}

Convolution in deep learning (in math class it is cross-correlation)

filter

w_{00}	w_{01}	w_{02}
w_{10}	w_{11}	w_{12}
w_{20}	w_{21}	w_{22}

Image

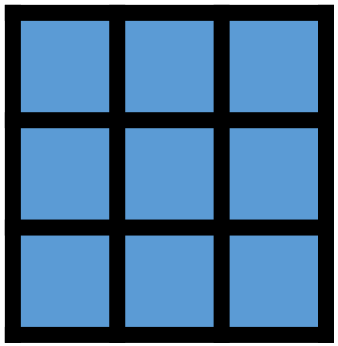
a_{00}	a_{01}	a_{02}	a_{03}	a_{04}	a_{05}
a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
a_{20}	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}
a_{30}	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}
a_{40}	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}
a_{50}	a_{51}	a_{52}	a_{53}	a_{54}	a_{55}

Note: Image-processing/math convolution is slightly different.
The kernel/filter is flipped around both axises before the multiplication.

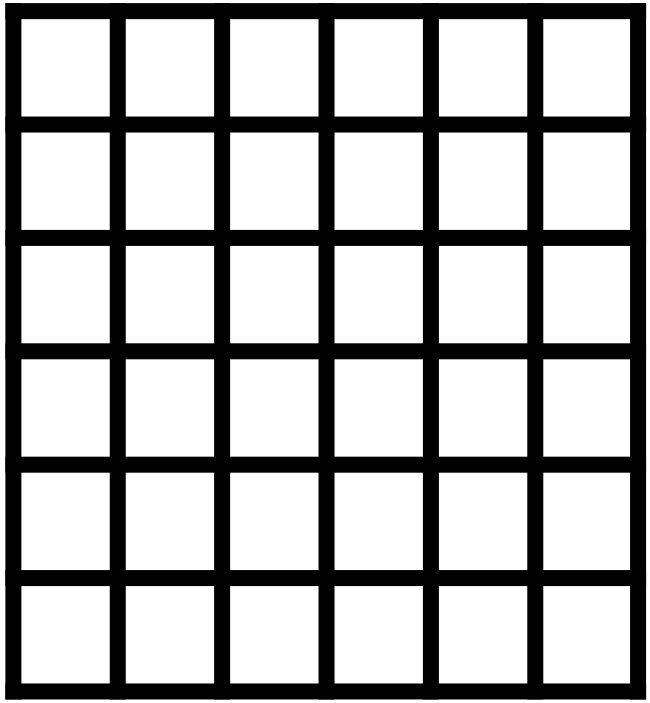
$$a'_{11} = w_{00} \cdot a_{00} + w_{01} \cdot a_{01} + w_{02} \cdot a_{02} + w_{10} \cdot a_{10} + w_{11} \cdot a_{11} + w_{12} \cdot a_{12} + w_{20} \cdot a_{20} + w_{21} \cdot a_{21} + w_{22} \cdot a_{22}$$

Convolution for images

filter

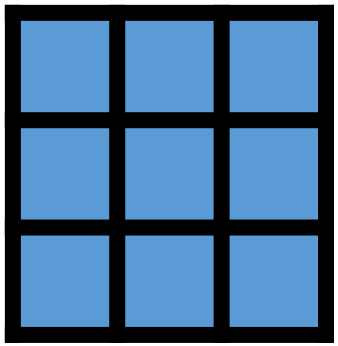


Image

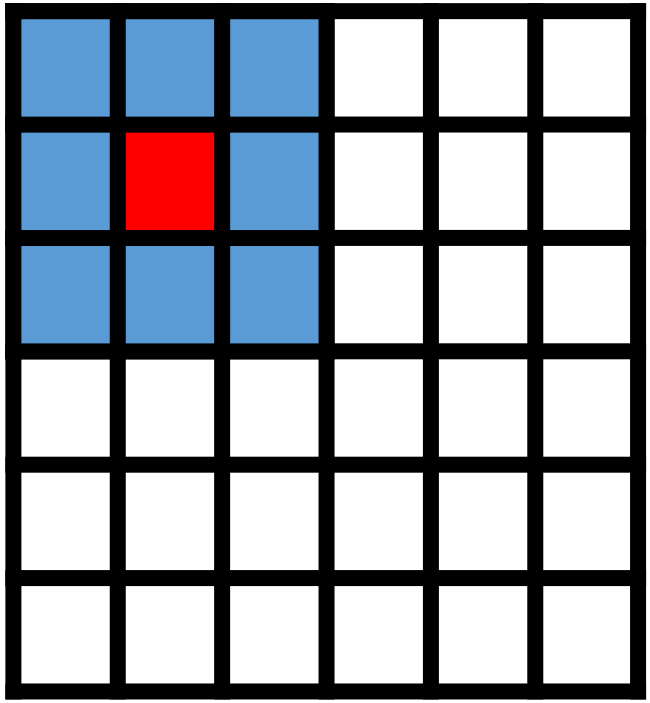


Convolution for images

filter

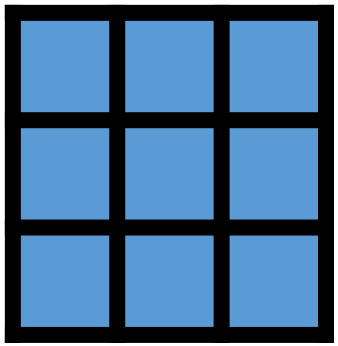


Image

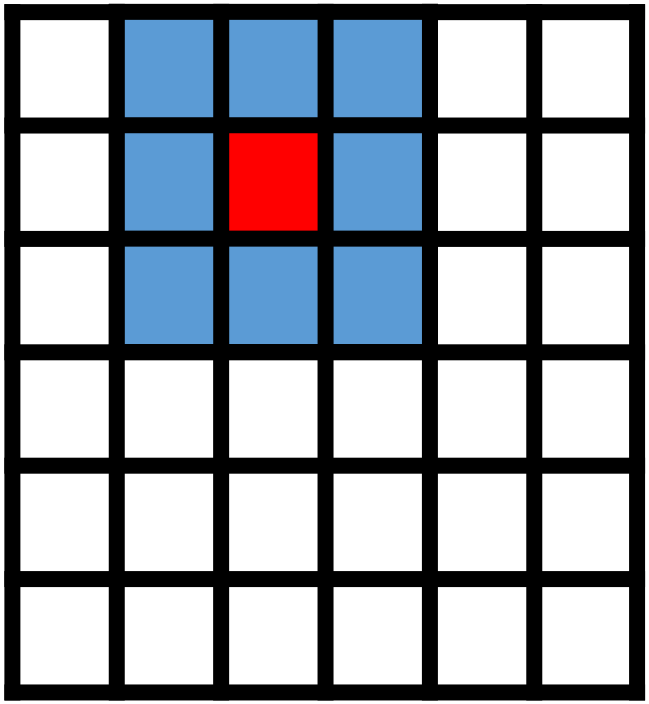


Convolution for images

filter

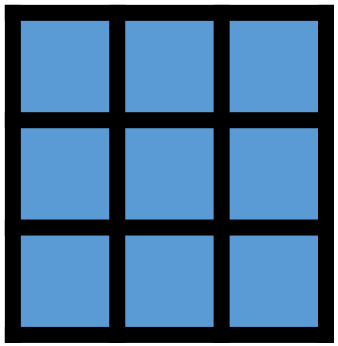


Image

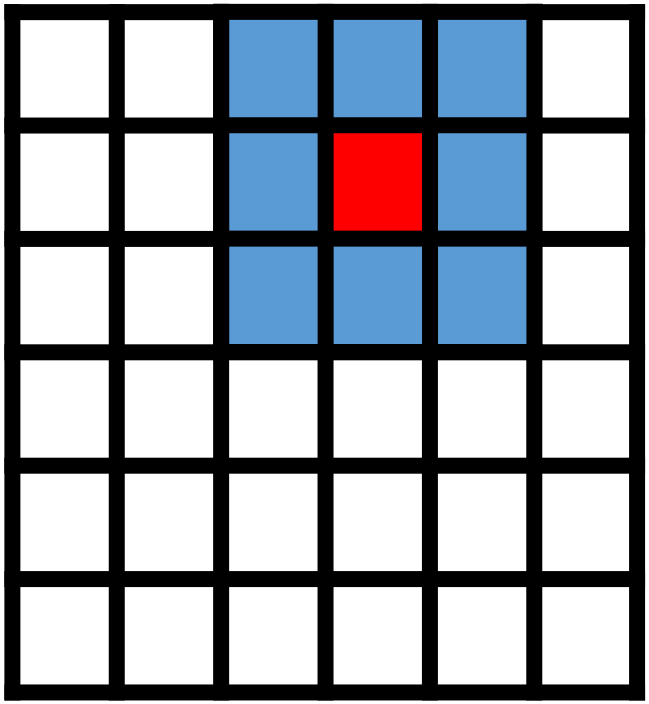


Convolution for images

filter

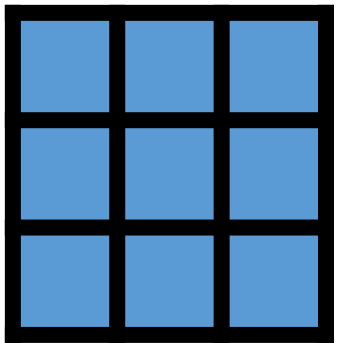


Image

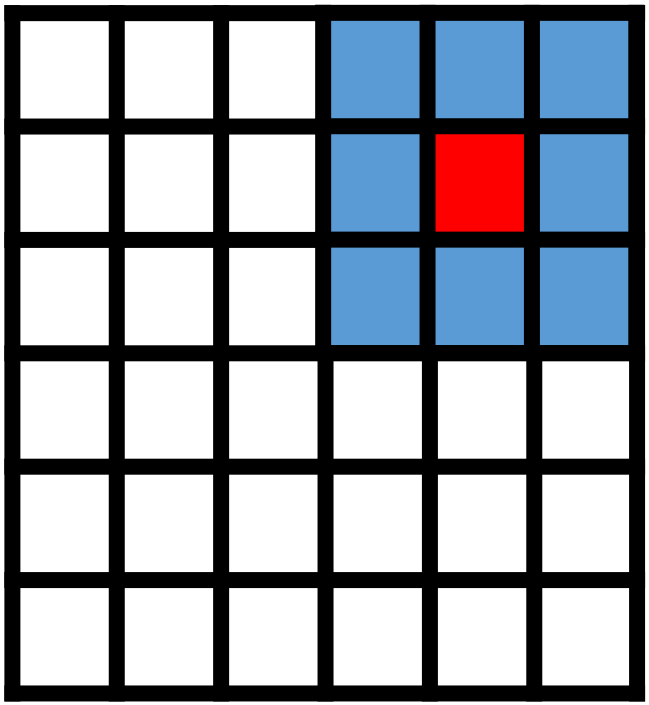


Convolution for images

filter

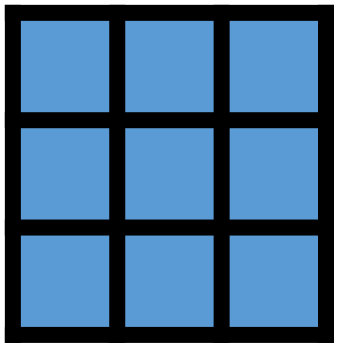


Image

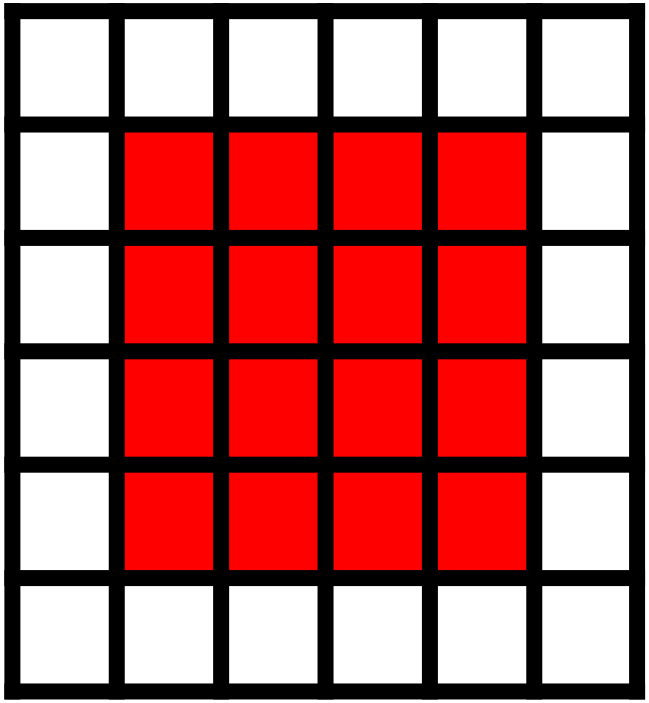


Convolution for images

filter

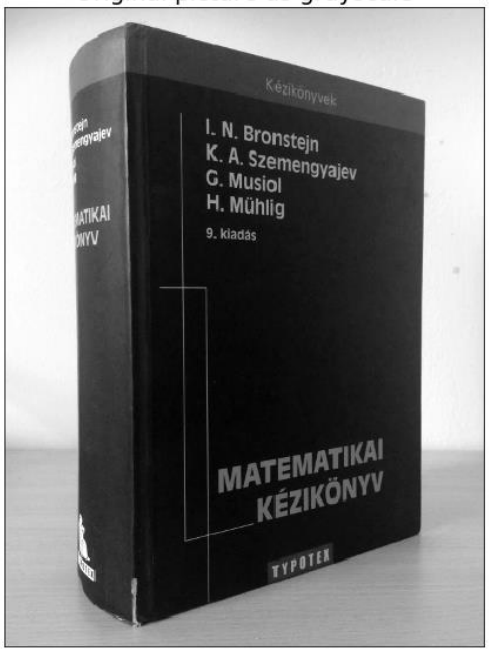


Image



Convolution for images - examples

Original picture as grayscale



*

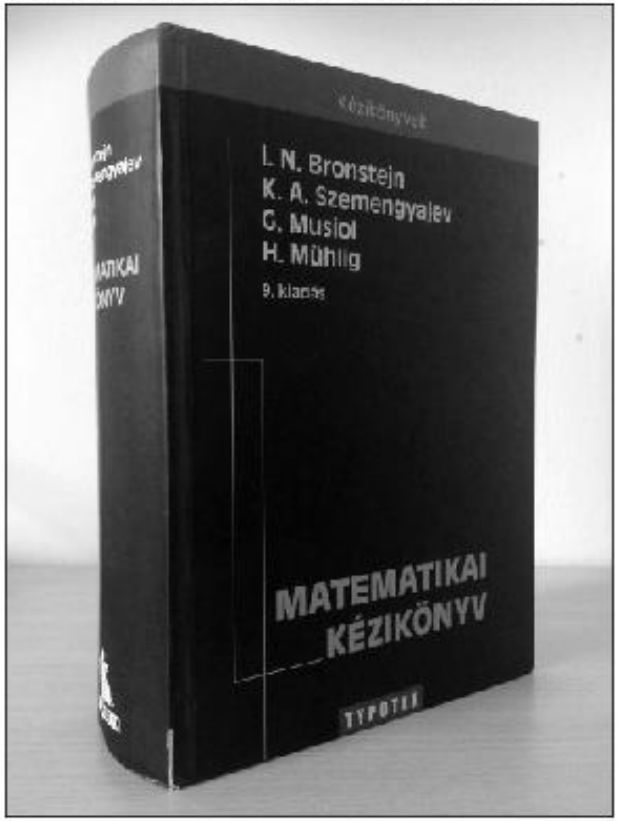
filter

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

Question: what do you expect?

Convolution for images - examples

Original picture as grayscale



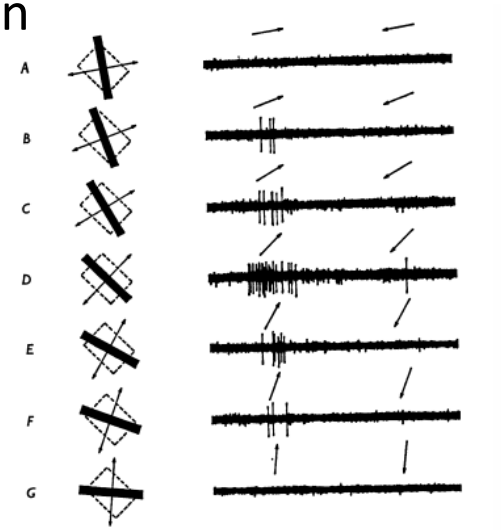
$$* \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} =$$

Horizontal edges

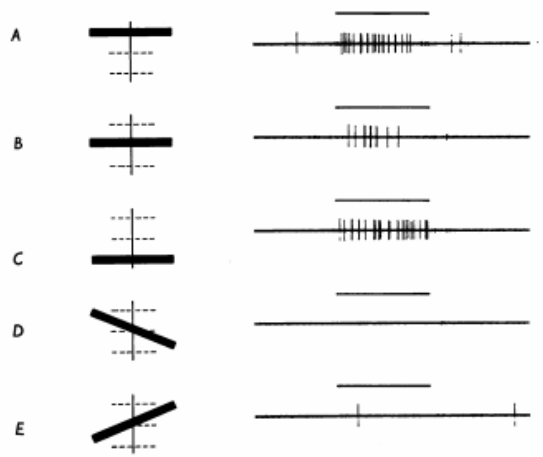
Excellent visualisations:
http://setosa.io/ev/image-kernels/
https://github.com/vdumoulin/conv_arithmetic

Neuroscientific experiments

- Hubel, Wiesel, Sperry
- cat's &monkey's vision
- Electrodes to the brain
- 1981 Nobel prize



Text-fig. 2. Responses of a complex cell in right striate cortex (layer IV A) t various orientations of a moving black bar. Receptive field in the left eye indicate by the interrupted rectangles; it was approximately $\frac{3}{4} \times \frac{3}{4}°$ in size, and was situate 4° below and to the left of the point of fixation. Ocular-dominance group 4. Duratio of each record, 2 sec. Background intensity 1·3 $\log_{10}$ cd/m², dark bars 0·0 log cd/m

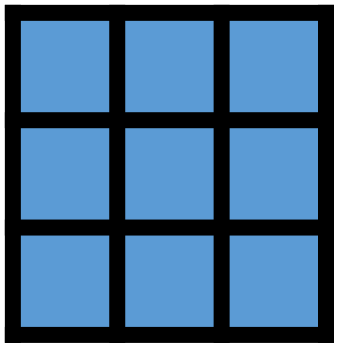


Text-fig. 7. Cell activated only by left (contralateral) eye over a field approximately $5 \times 5°$, situated 10° above and to the left of the area centralis. The cell responded best to a black horizontal rectangle, $\frac{1}{2} \times 6°$, placed anywhere in the receptive field (*A–C*). Tilting the stimulus rendered it ineffective (*D–E*). The black bar was introduced against a light background during periods of 1 sec, indicated by the upper line in each record. Luminance of white background, 1·0 $\log_{10}$ cd/m²; luminance of black part, 0·0 $\log_{10}$ cd/m². A lesion, made while recording from the cell, was found in layer 2 of apical segment of post-lateral gyrus.

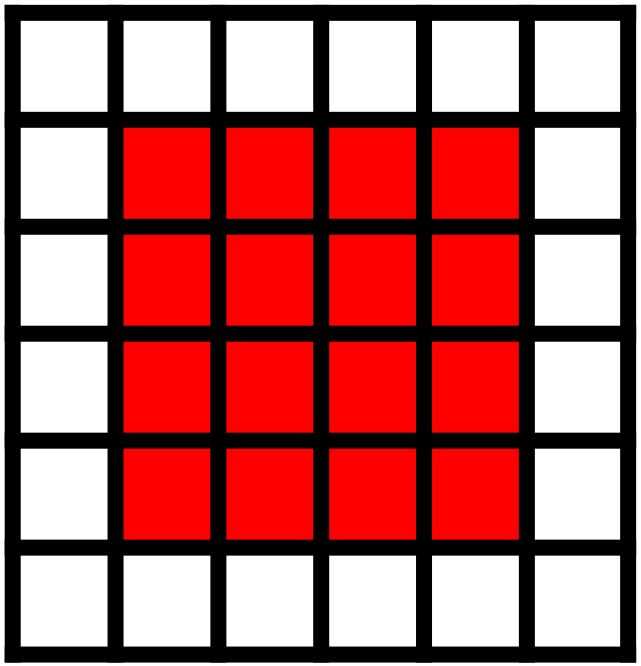
[Hubel, Wiesel: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, 1961]
[Hubel, Wiesel: Receptive fields and functional architecture of monkey striate cortex, 1968]

Padding

filter

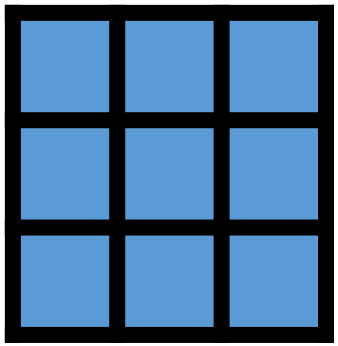


Image



Padding: 1

filter

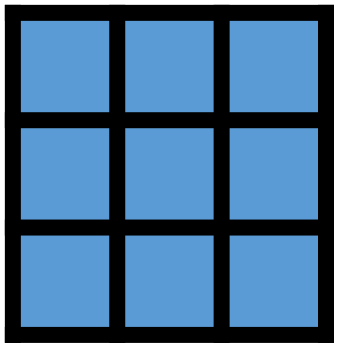


Image

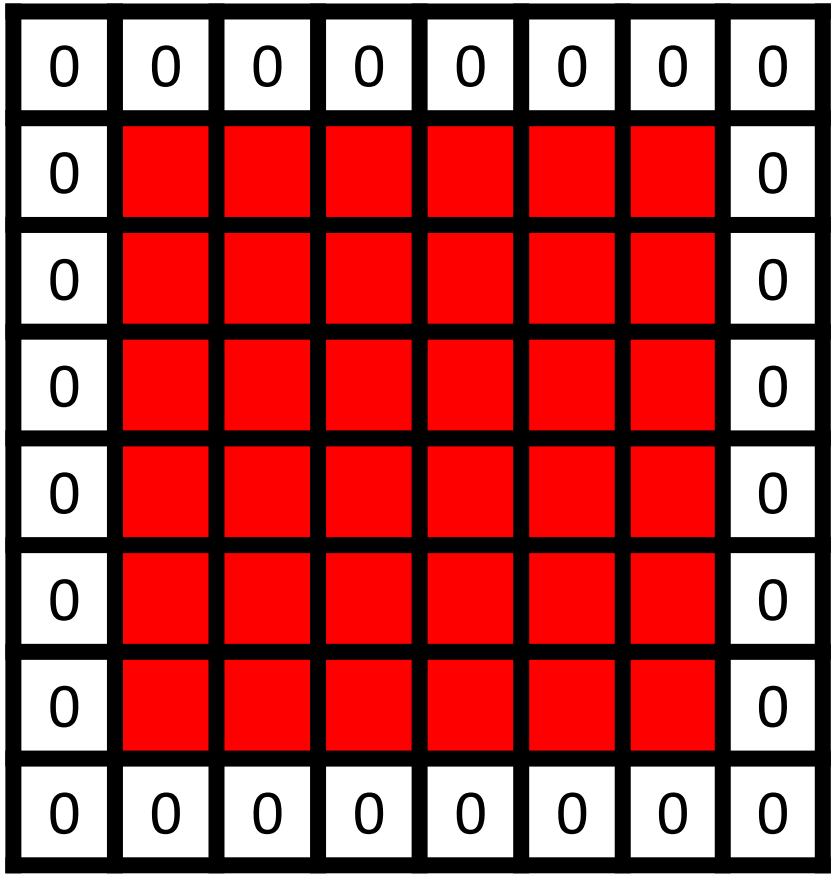
0	0	0	0	0	0	0	0
0							0
0							0
0							0
0							0
0							0
0							0
0	0	0	0	0	0	0	0

Padding: 1

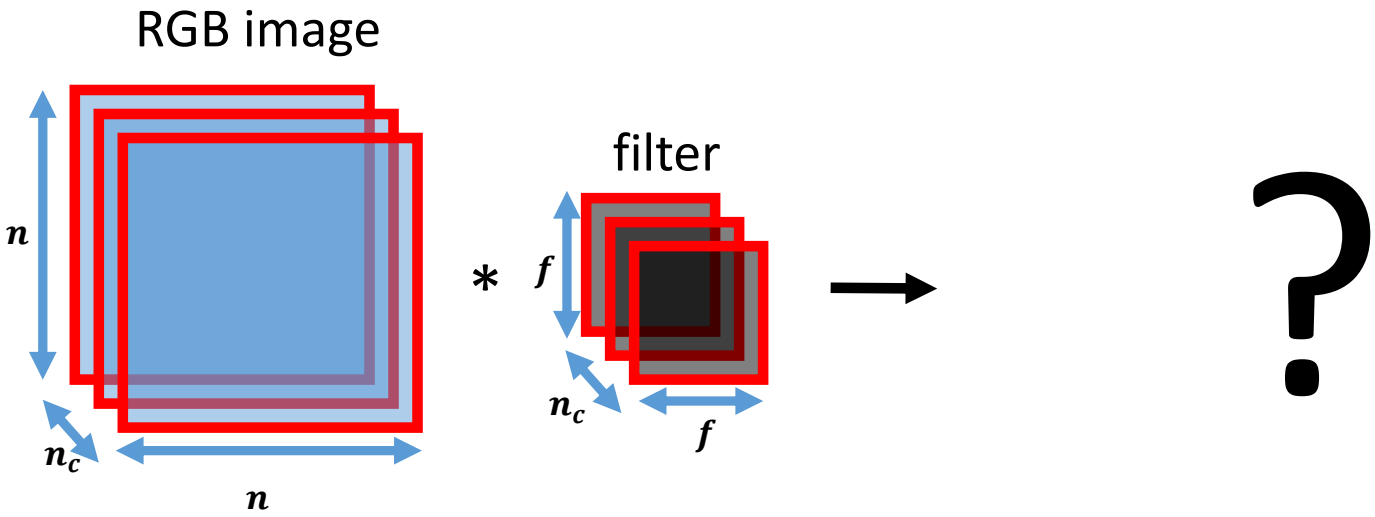
filter



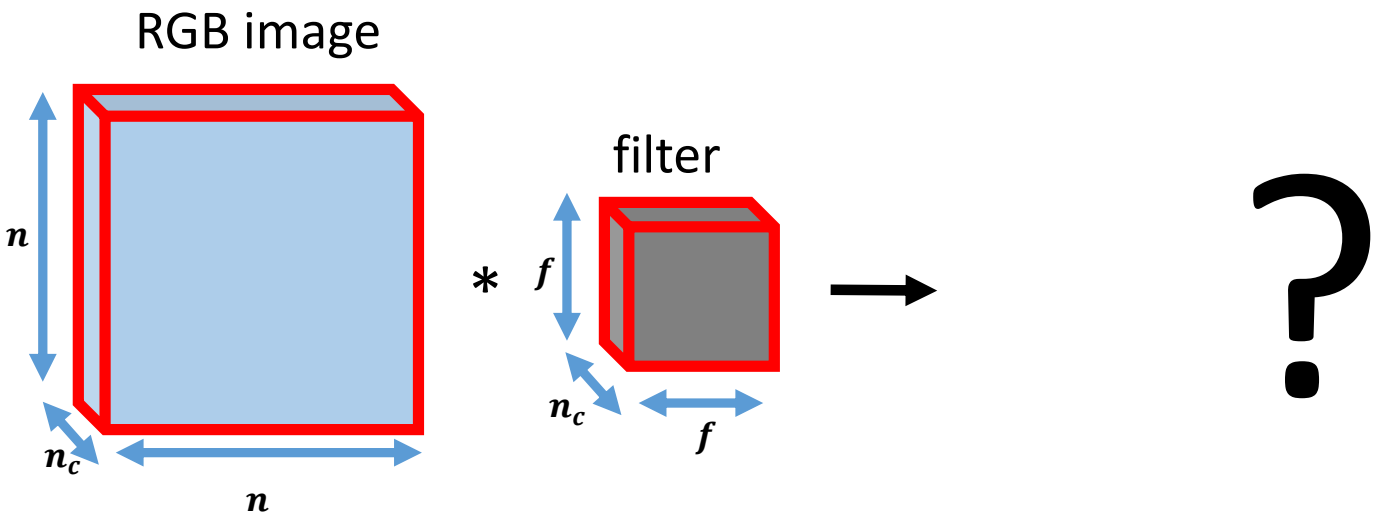
Image



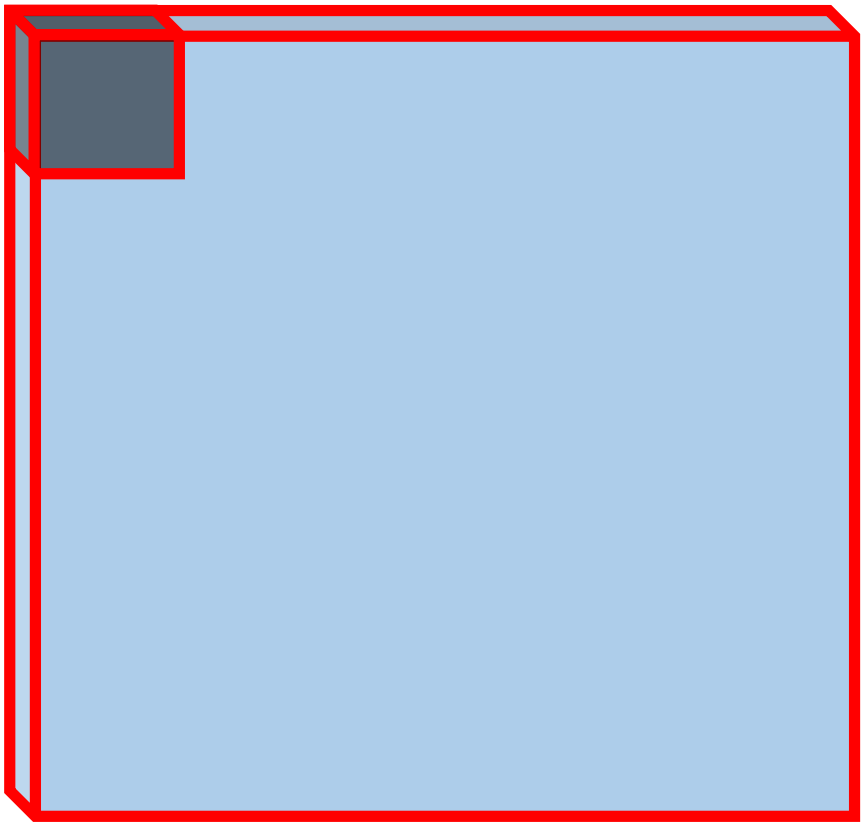
Convolution over volume



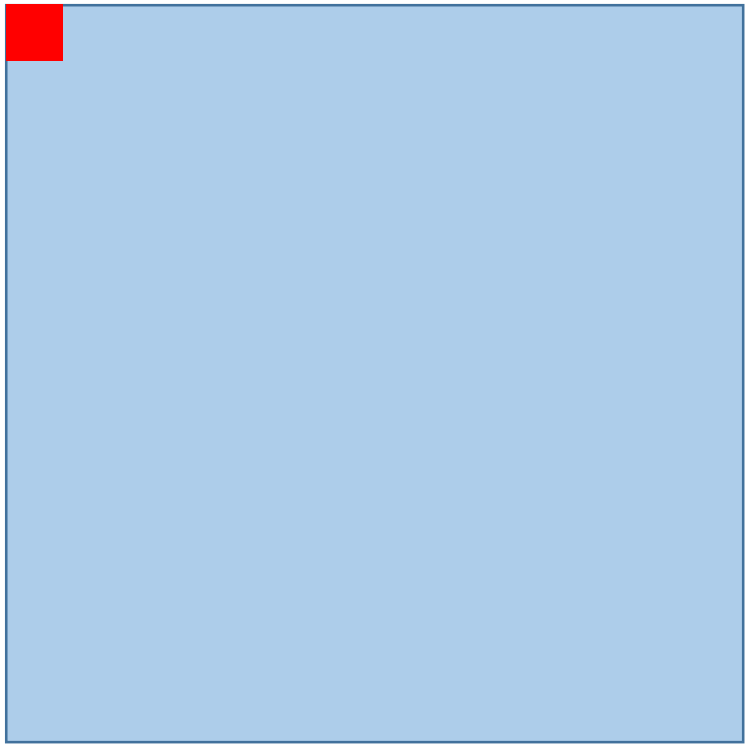
Convolution over volume



Convolution over volume

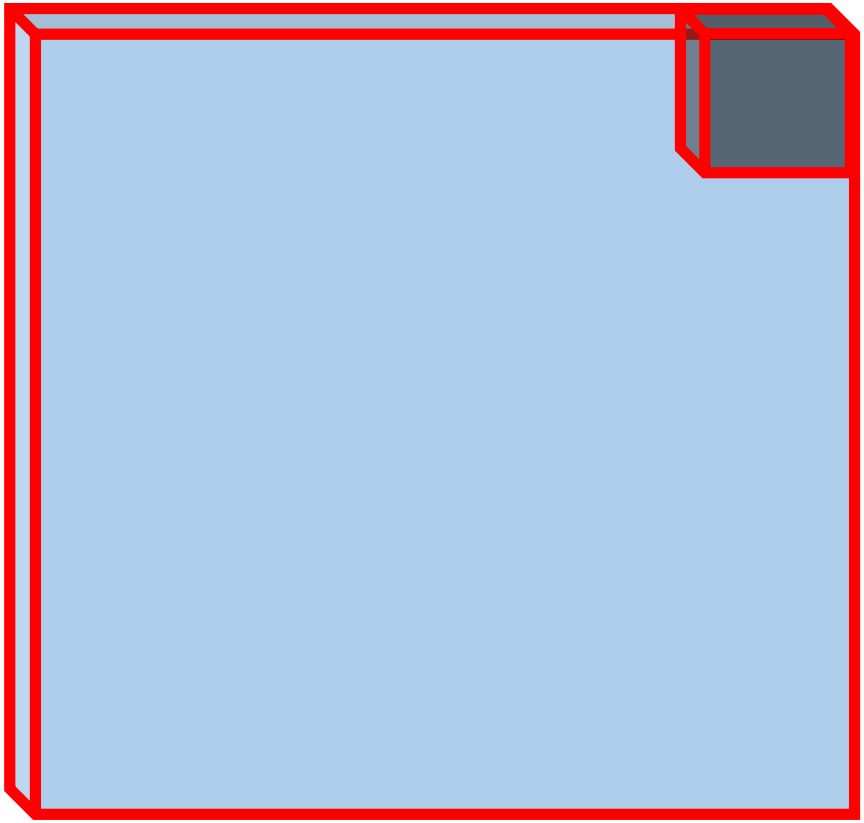


n_c = 3 (RGB)

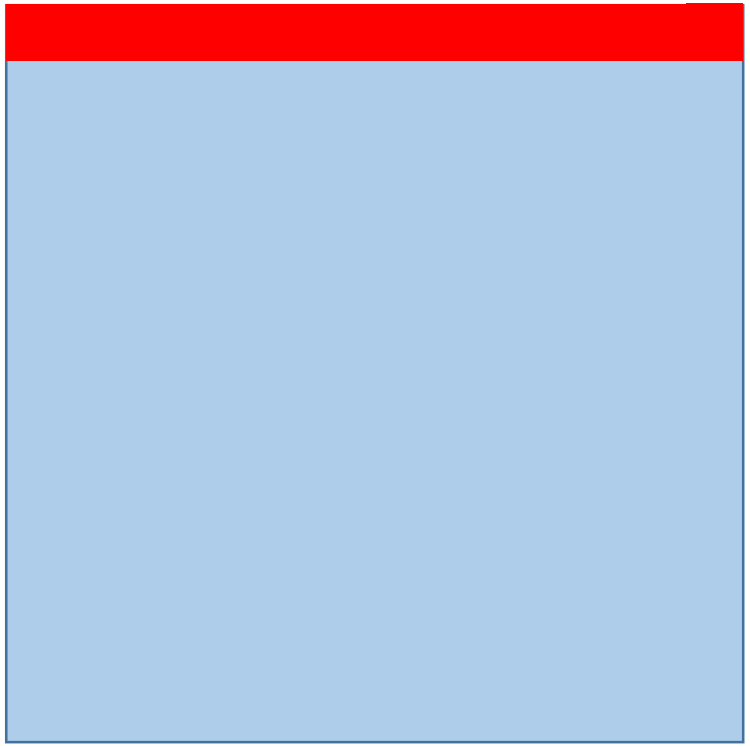


n_c = 1

Convolution over volume

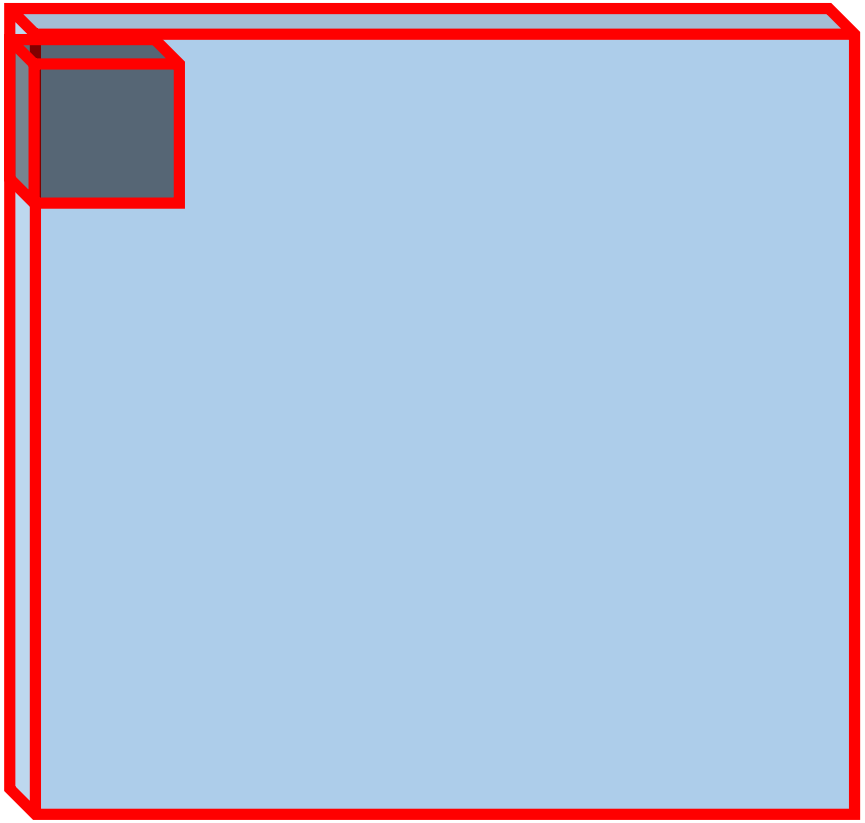


n_c = 3 (RGB)

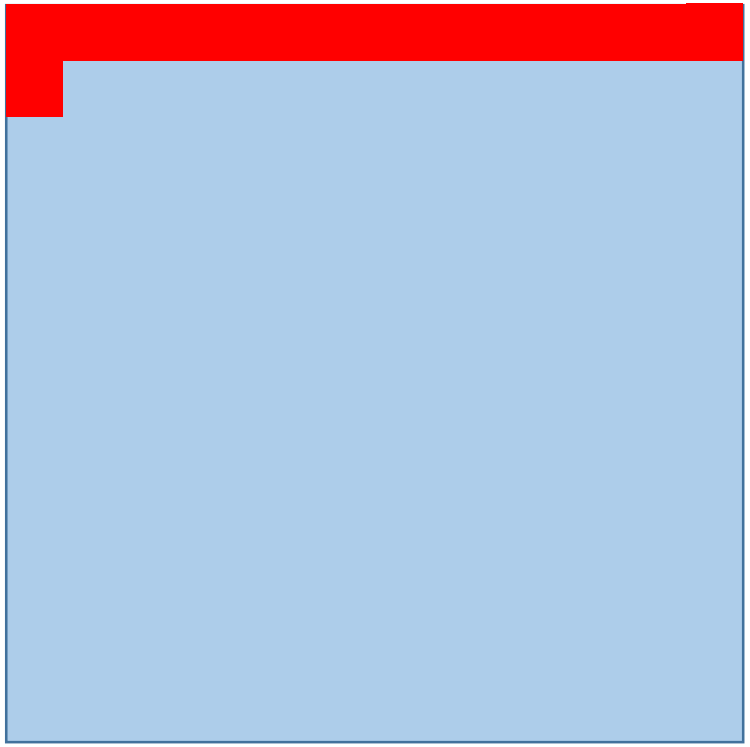


n_c = 1

Convolution over volume

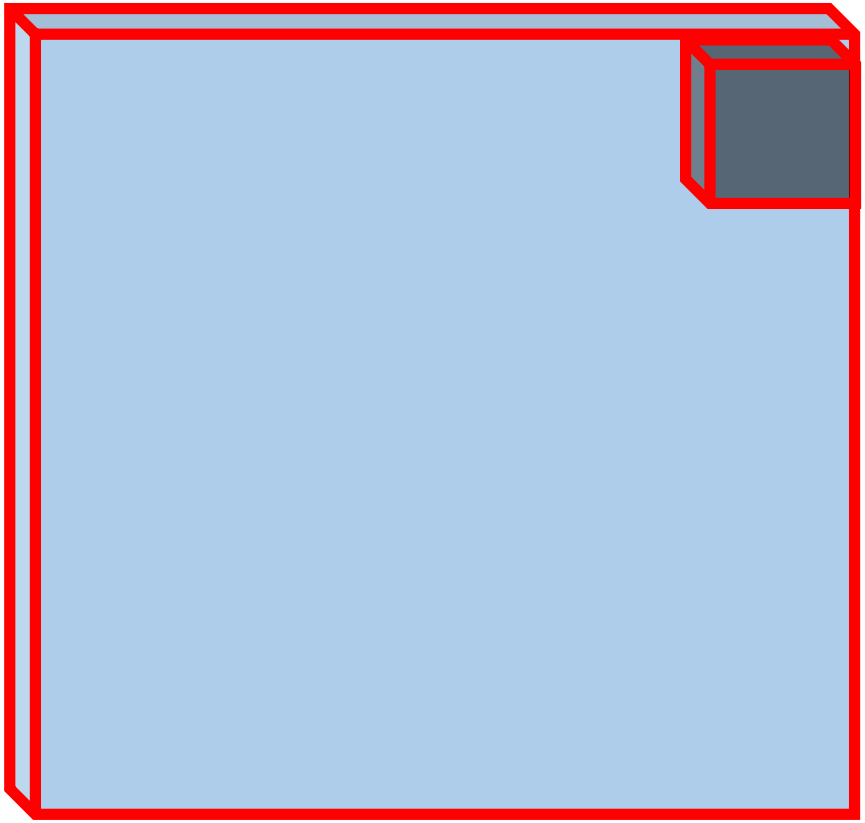


n_c = 3 (RGB)

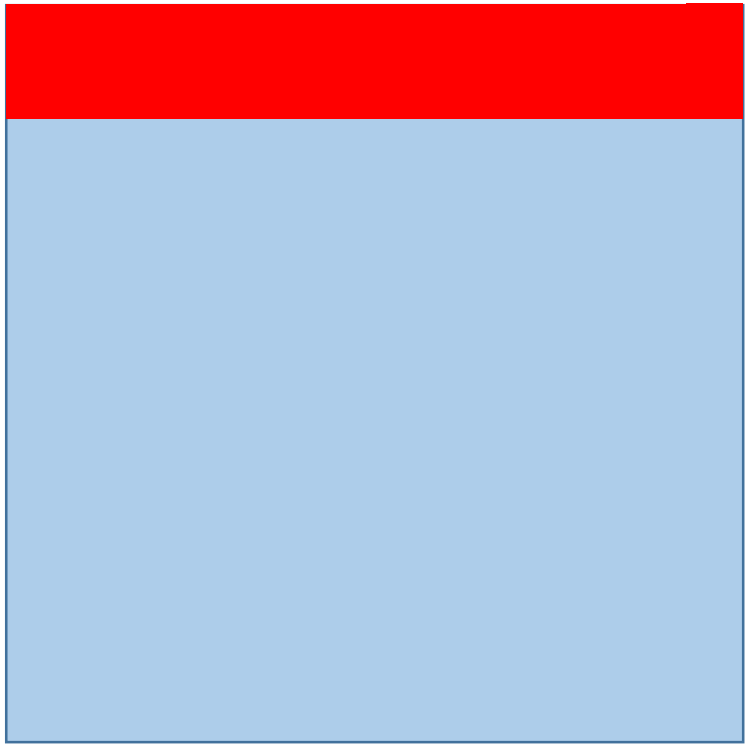


n_c = 1

Convolution over volume

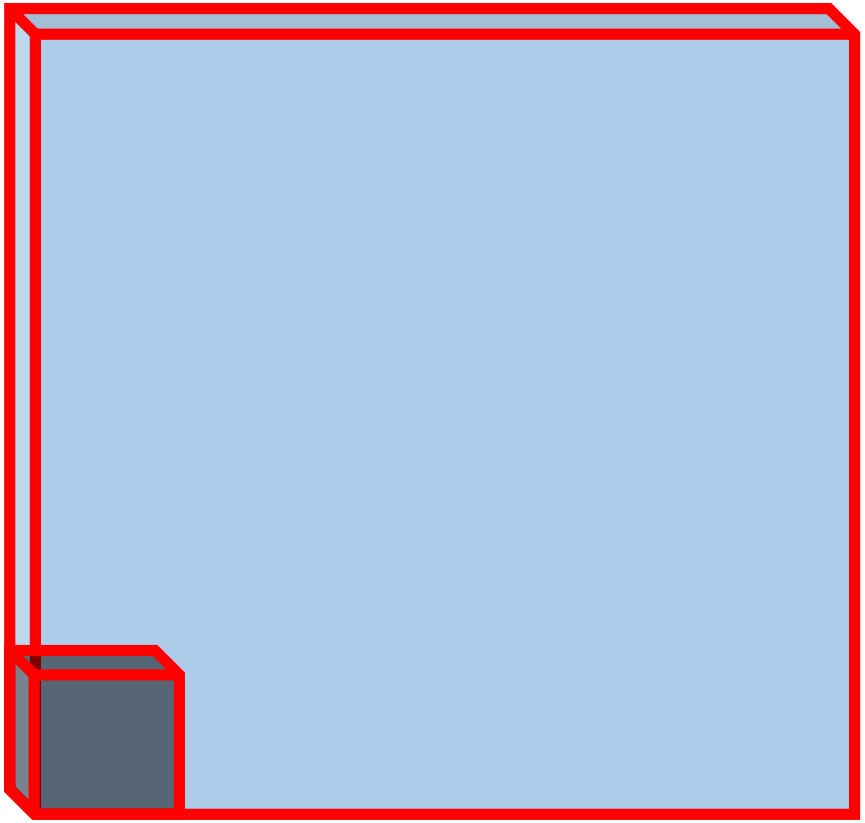


n_c = 3 (RGB)

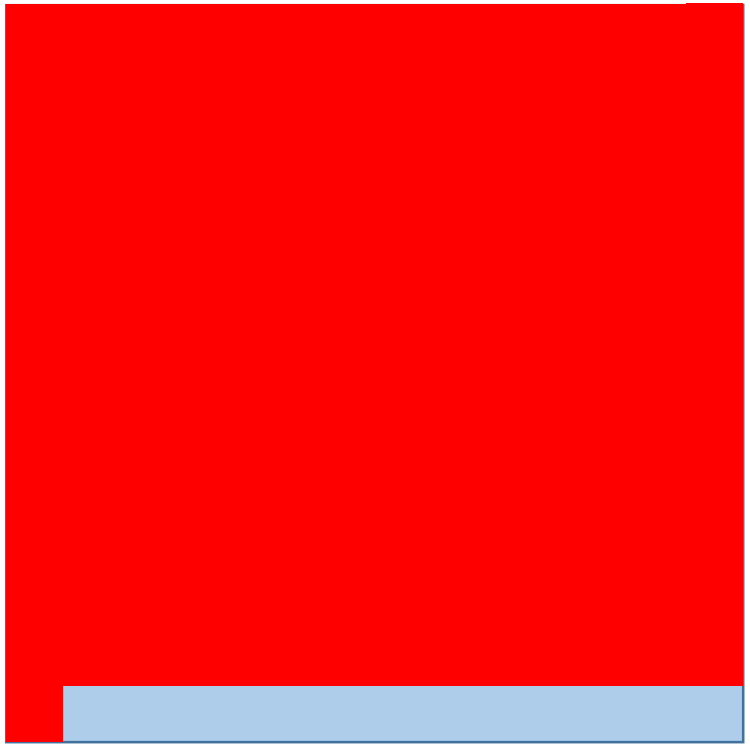


n_c = 1

Convolution over volume

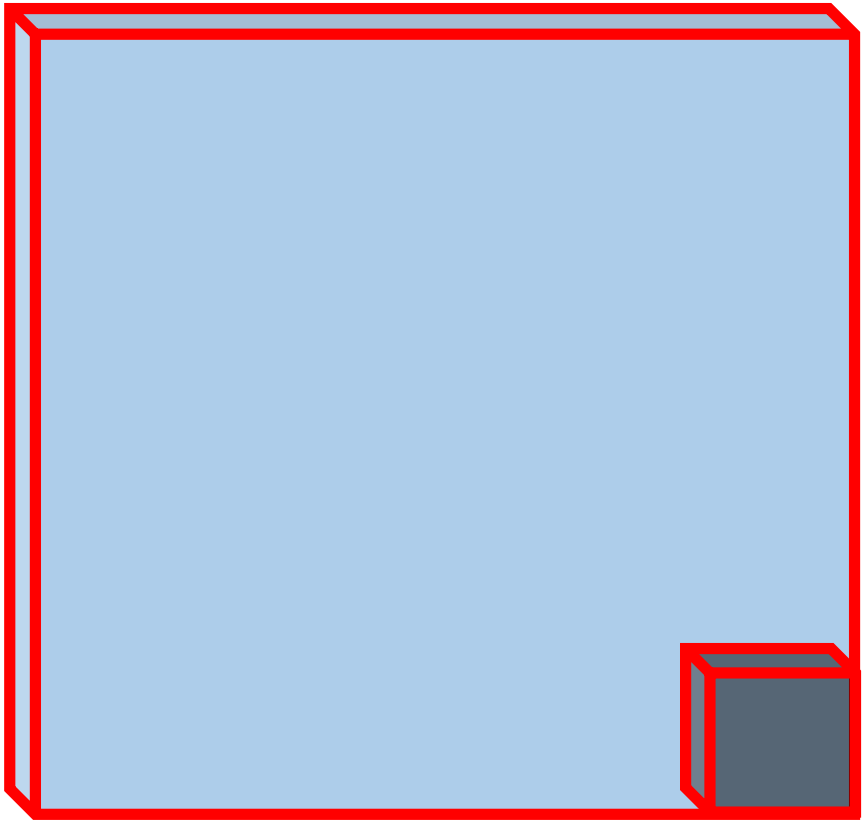


n_c = 3 (RGB)

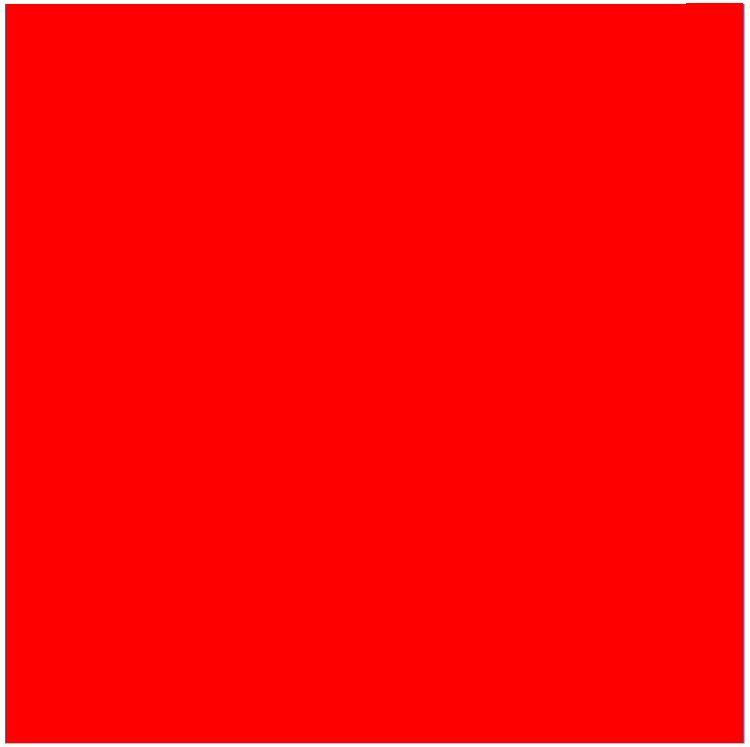


n_c = 1

Convolution over volume

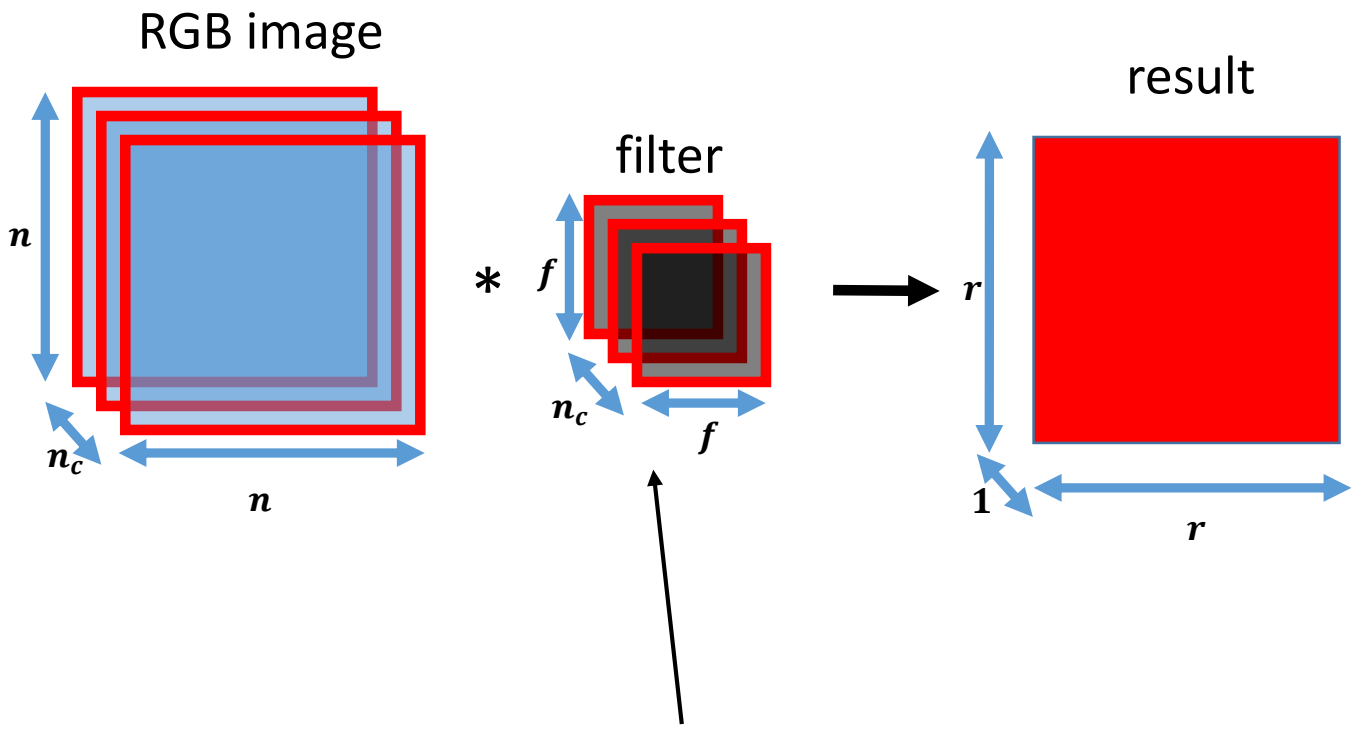


n_c = 3 (RGB)



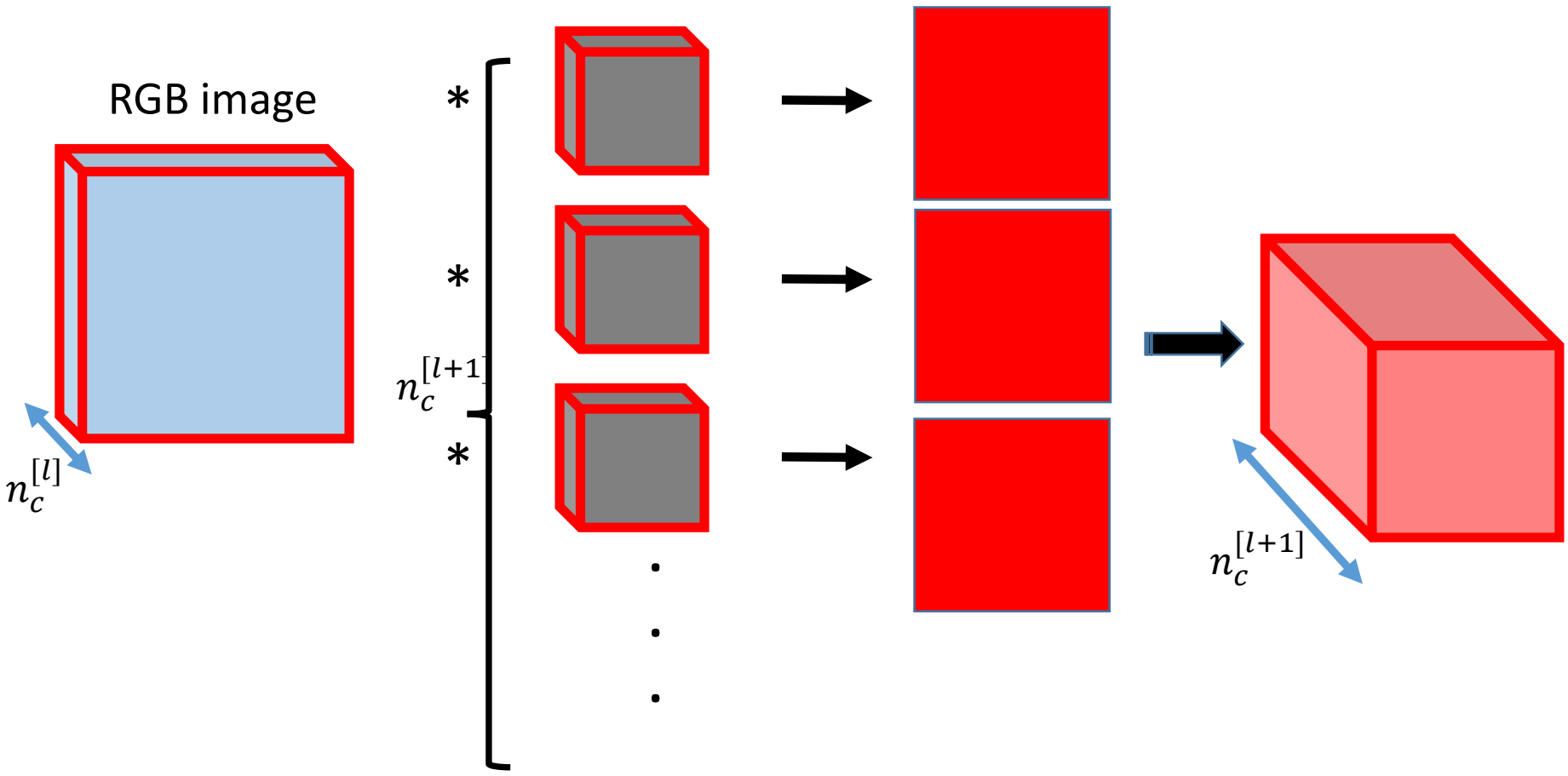
n_c = 1

Convolution over volume



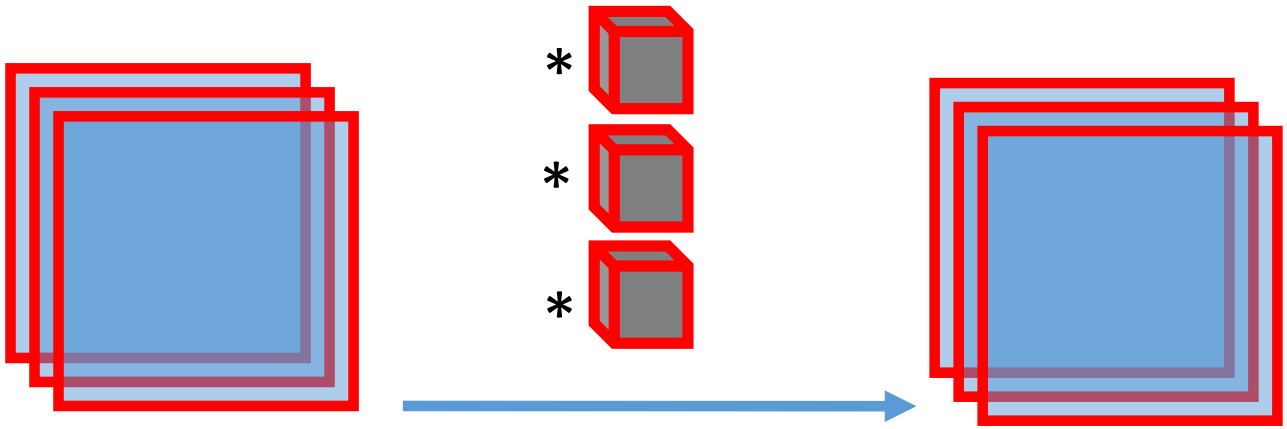
these are n_c independent filters each with different parameters

Convolution over volume



1 layer of convolution in neural networks

$\#channels\ in\ layer\ l: n_c^{[l]}$



$n_c^{[0]} = 3$ $n_c^{[1]} = 3$

Difference from the previous: 1 bias per filter

Each filter has a parameter number of: $f \cdot f \cdot n_c + 1$

1 layer of convolution in neural networks

Input: 200 x 200 pixel RGB image, in the first layer we want 200 x 200 x 3 neurons

How many parameters does it have?

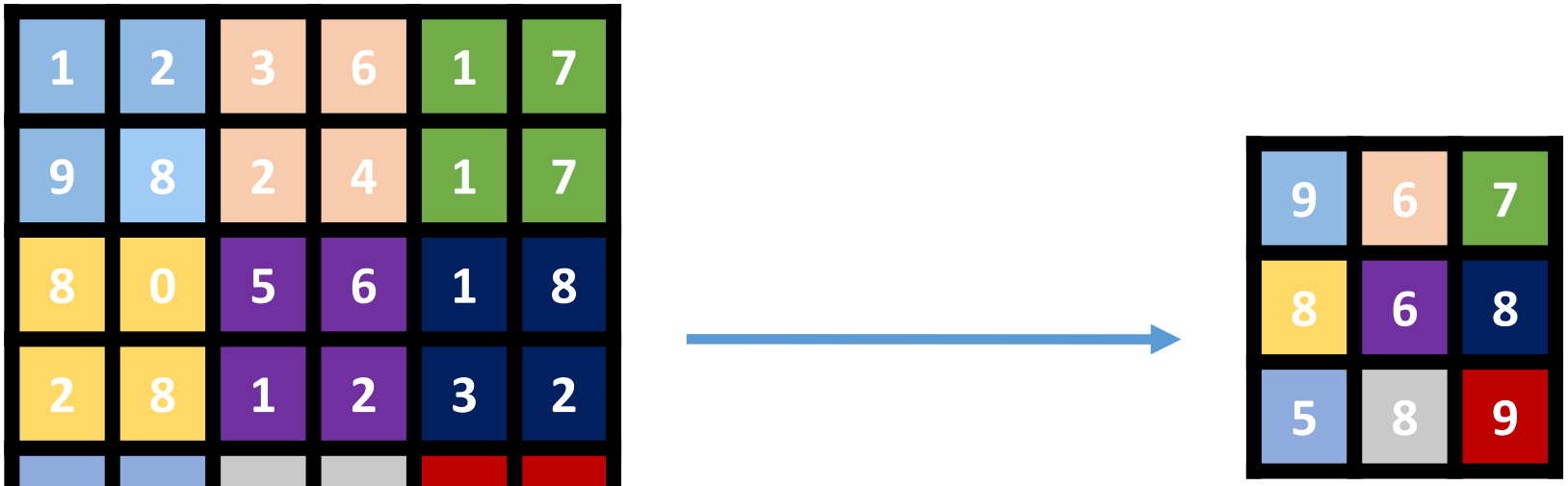
Fully connected layer

- weights: $(200 \cdot 200 \cdot 3)^2$
- bias: $200 \cdot 200 \cdot 3$
- $Total \approx 1.4 \cdot 10^{10}$

Convolutional layer (f = 3)

- Weights per filter: $f \cdot f \cdot n_c^{[0]} + 1$
- Number of filters: $n_c^{[1]}$
- $Total = 84$

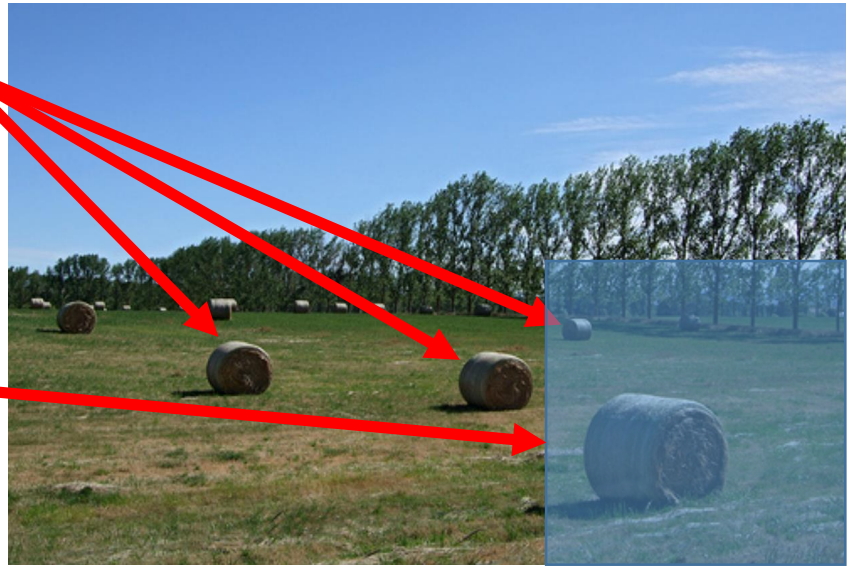
Maxpooling: a special filter to reduce parameters



Translation invariance

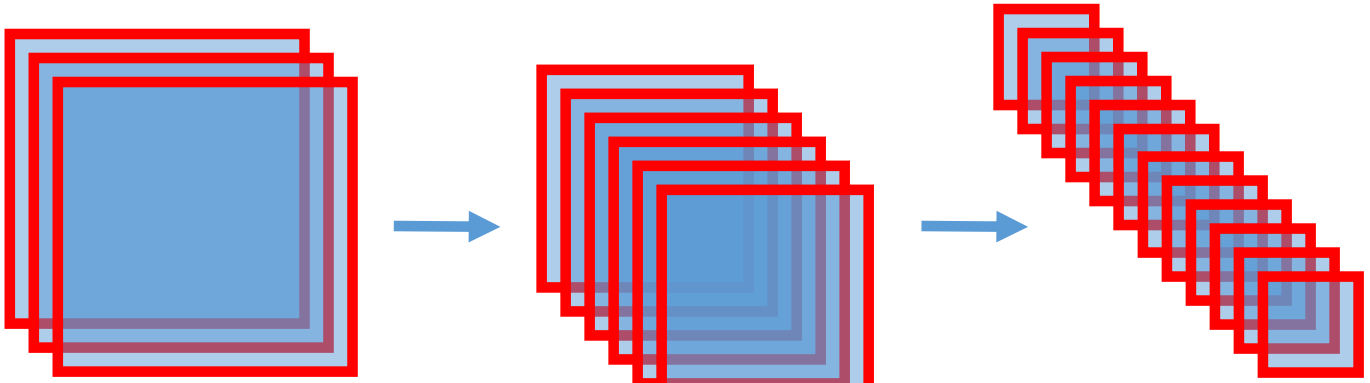
- Objects are not position dependents

✓ A convolution filter
✓ Maxpooling: best value from a region (exact position doesn't matter for image classification)



Convolution in neural networks - representation

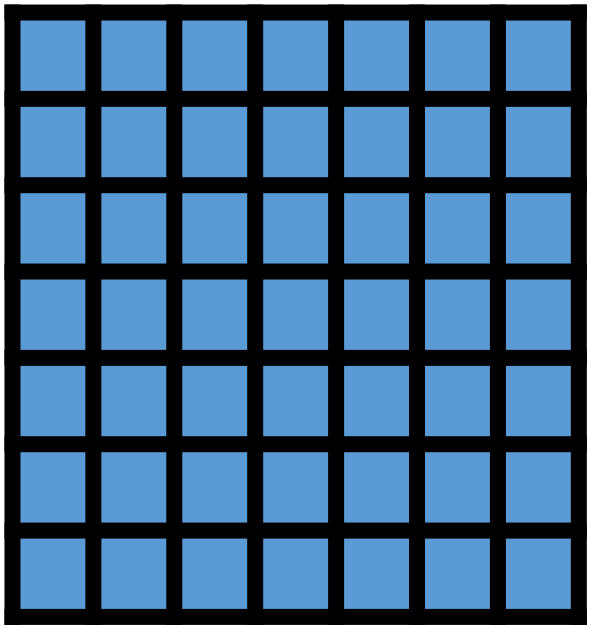
$\#channels\ in\ layer\ l: n_c^{[l]}$



$n_c^{[0]} = 3$ $n_c^{[1]} = 6$ $n_c^{[2]} = 12$

Small f vs large f – receptive field

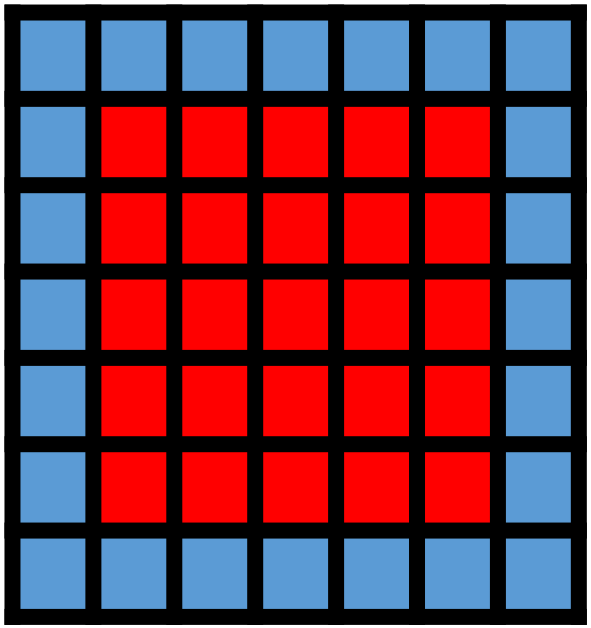
Three 3 x 3 convolution after each other
No padding, stride=1, f=3



Small f vs large f – receptive field

Three 3 x 3 convolution after each other
No padding, stride=1, f=3

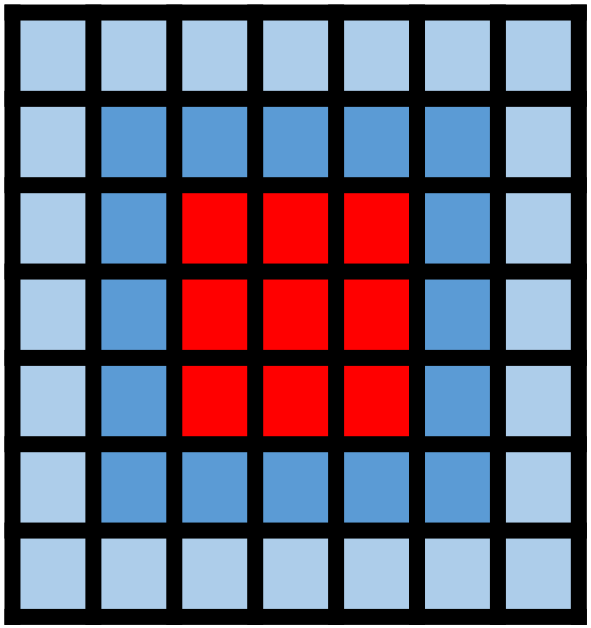
After 1 layer:



Small f vs large f – receptive field

Three 3 x 3 convolution after each other
No padding, stride=1, f=3

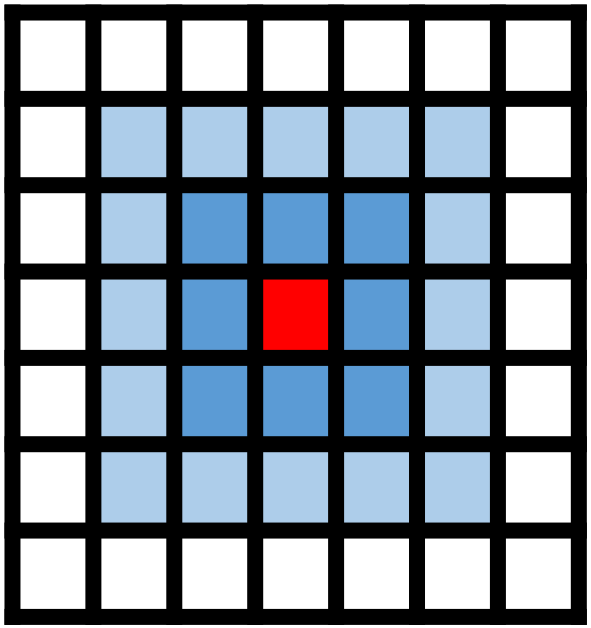
After 2 layer:



Small f vs large f – receptive field

Three 3 x 3 convolution after each other
No padding, stride=1, f=3

After 3 layer:



Small f vs large f – receptive field

After three 3 x 3 convolution each neuron can see 7 x 7 field from the input

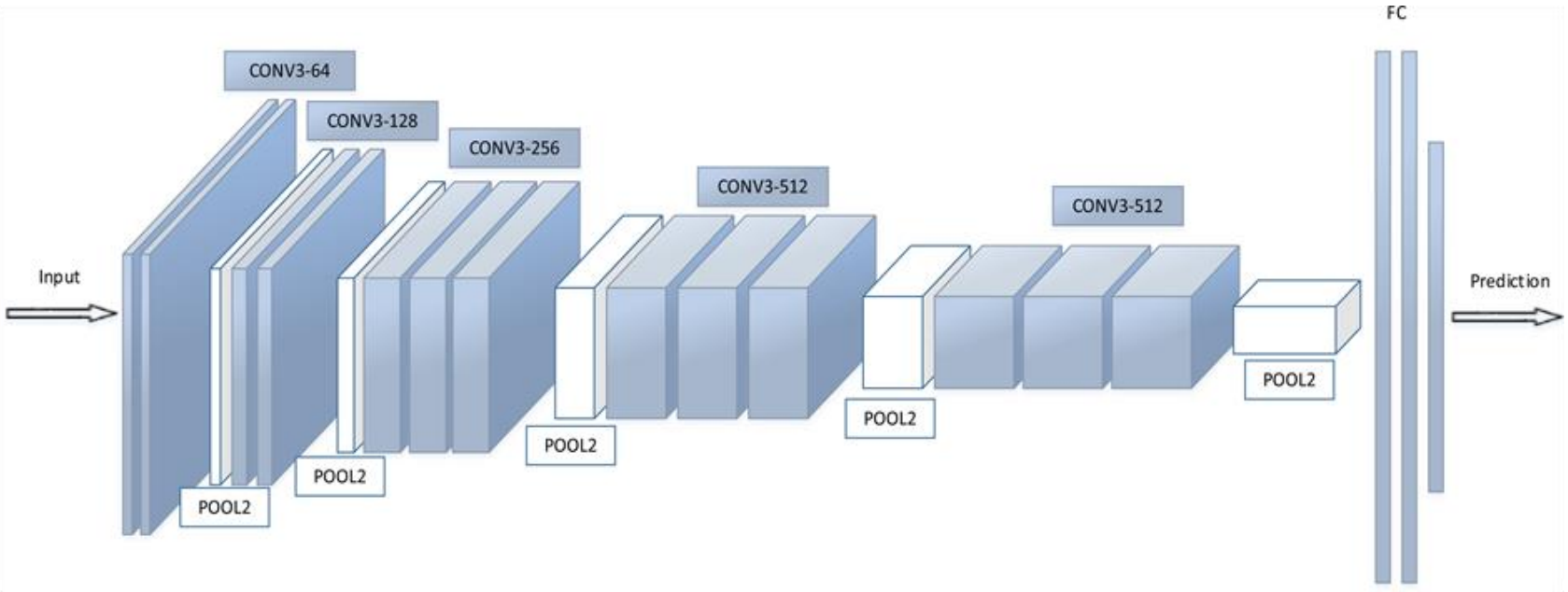
#parameters in three 3 x 3 conv: $3 \cdot (3 \cdot 3 + 1) = 30$

#parameters in one 7 x 7 conv: $1 \cdot (7 \cdot 7 + 1) = 50$

3 convolutions → more 'non-linearity'

VGG16

This is actually a smart way to restrict our universal function.



[http://file.scirp.org/Html/4-7800353_65406.htm]

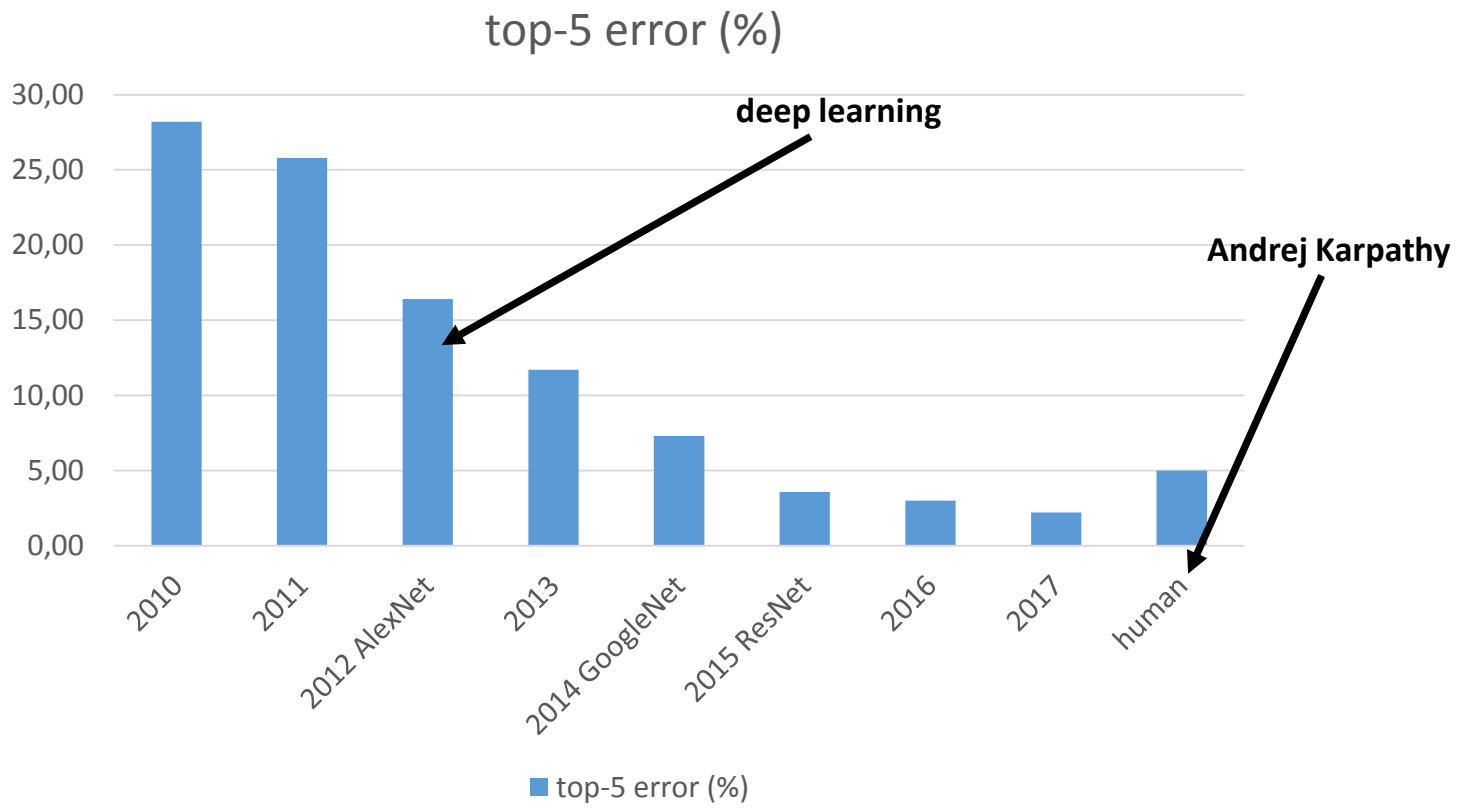
Nice visualisation of CNNs:
http://scs.ryerson.ca/~aharley/vis/conv/flat.html

https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/

ImageNet Large Scale Visual Recognition Challange

- 2010 –
- 1.2M images (100K test set)
- 1000 categories
- 'Image classification world cup',
- top-5 error (still not that easy…)

ImageNet Large Scale Visual Recognition Challange



What about you?

https://cs.stanford.edu/people/karpathy/ilsvrc/

Keras CNN notebook

https://github.com/patbaa/physdl/blob/master/notebooks/05/cnn.ipynb