

Современные нейросетевые технологии

Лекция 5. Свёрточные сети

1. Операции свертки 1D, 2D, 3D
2. Сверточные слои
3. Pooling слой

github.com/balezz/modern_dl
Срок сдачи А4 – 01.10.2022 г.

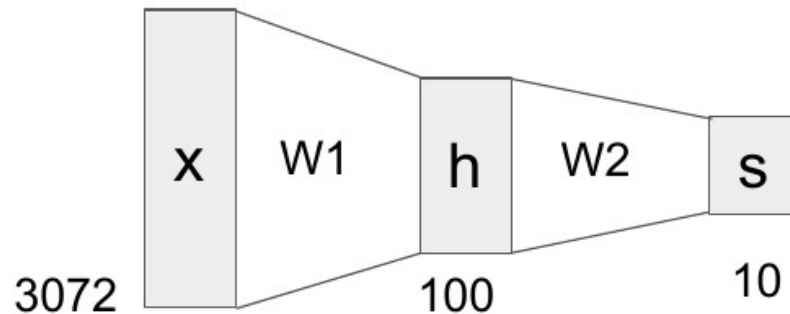
Источники:

- dlcourse.ai
- cs231n.stanford.edu
- cs230.stanford.edu

Neural networks: without the brain stuff

(Before) Linear score function: $f = Wx$

(Now) 2-layer Neural Network $f = W_2 \max(0, W_1 x)$



СВЕРТОЧНЫЕ СЕТИ

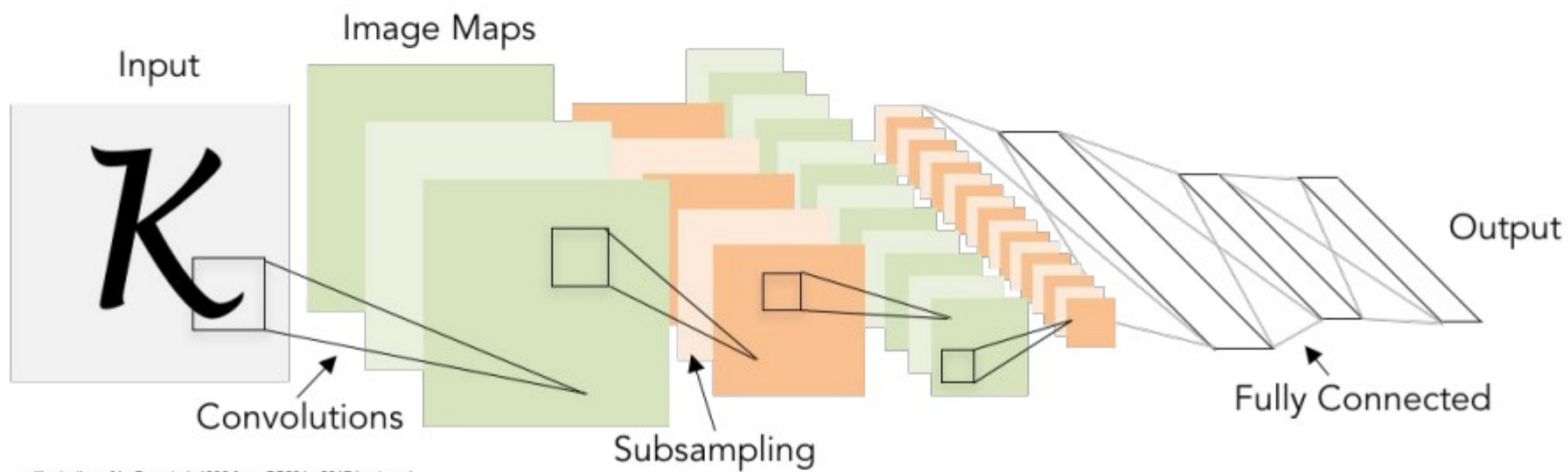


Illustration of LeCun et al. 1998 from CS231n 2017 Lecture 1

A bit of history:

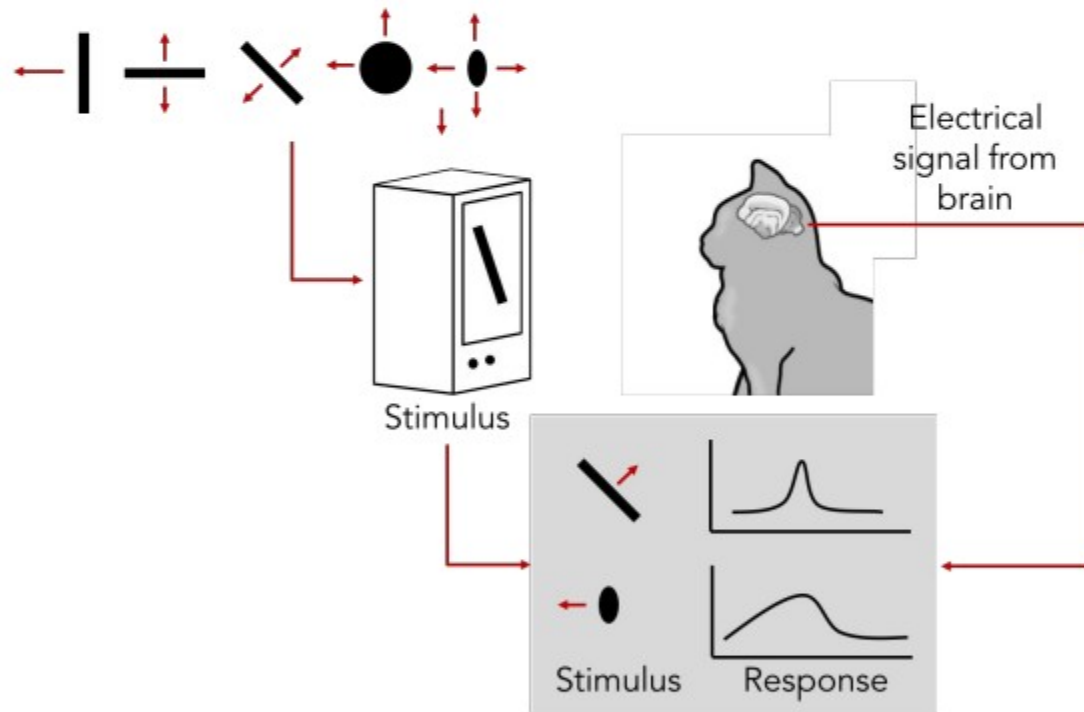
Hubel & Wiesel,
1959

RECEPTIVE FIELDS OF SINGLE
NEURONES IN
THE CAT'S STRIATE CORTEX

1962

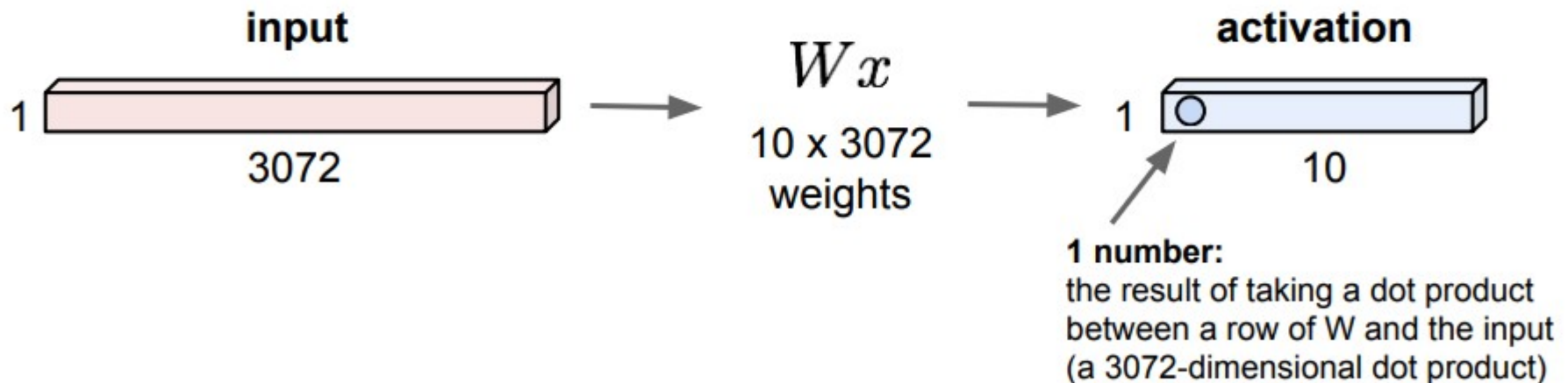
RECEPTIVE FIELDS, BINOCULAR
INTERACTION
AND FUNCTIONAL ARCHITECTURE IN
THE CAT'S VISUAL CORTEX

1968...



Fully Connected Layer

32x32x3 image -> stretch to 3072 x 1



1D СЕРПКА

[illegible]

$$c_i = \sum_{k=0}^{N_{out}} x_k w_{i-k}, \quad i = 0, \dots, N_{out}$$

2D СВЕРТКА

[0, 0, 1	1, 0, 0, 0]
[0, 0, 1	1, 0, 0, 0]
[0, 1, 1	1, 0, 0, 0]
[0, 0, 1, 1, 0	0, 0]
[0, 0, 1, 1, 0	0, 0]
[0, 0, 1, 1, 0	0, 0]

0, 0, 0
1, 1, 1
0, 0, 0

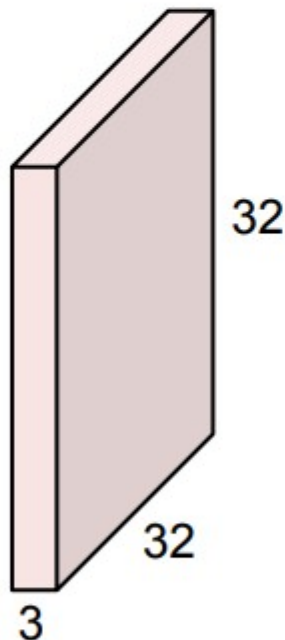
0, 1, 0
0, 1, 0
0, 1, 0

1.	2.	2.	1.	0.]
[2.	3.	2.	1.	0.]
[1.	2.	2.	1.	0.]
[1.	2.	2.	1.	0.]

[1.	3.	3.	0.	0.]
[1.	3.	3.	0.	0.]
[1.	3.	3.	0.	0.]
[0.	3.	3	0.	0.]

Convolution Layer

32x32x3 image



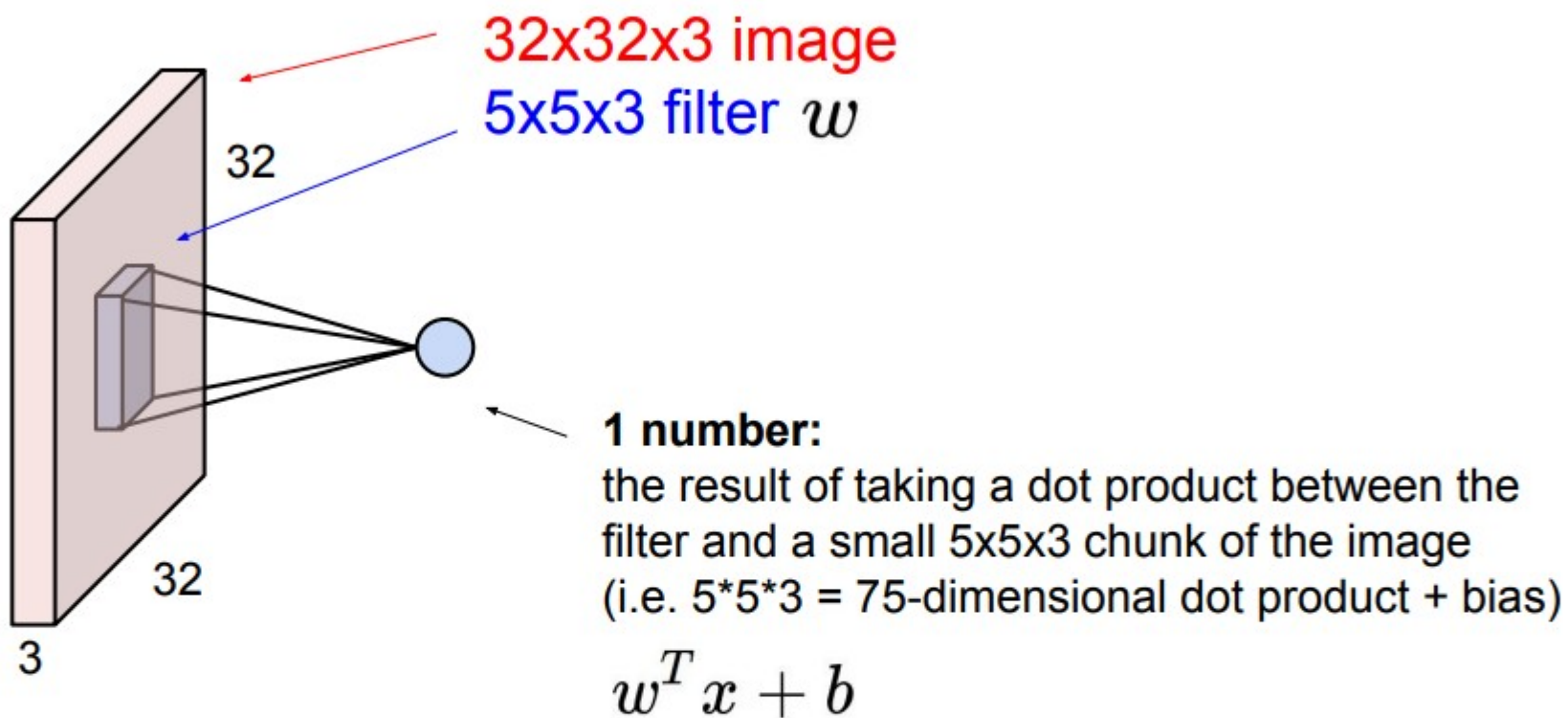
Filters always extend the full depth of the input volume

5x5x3 filter

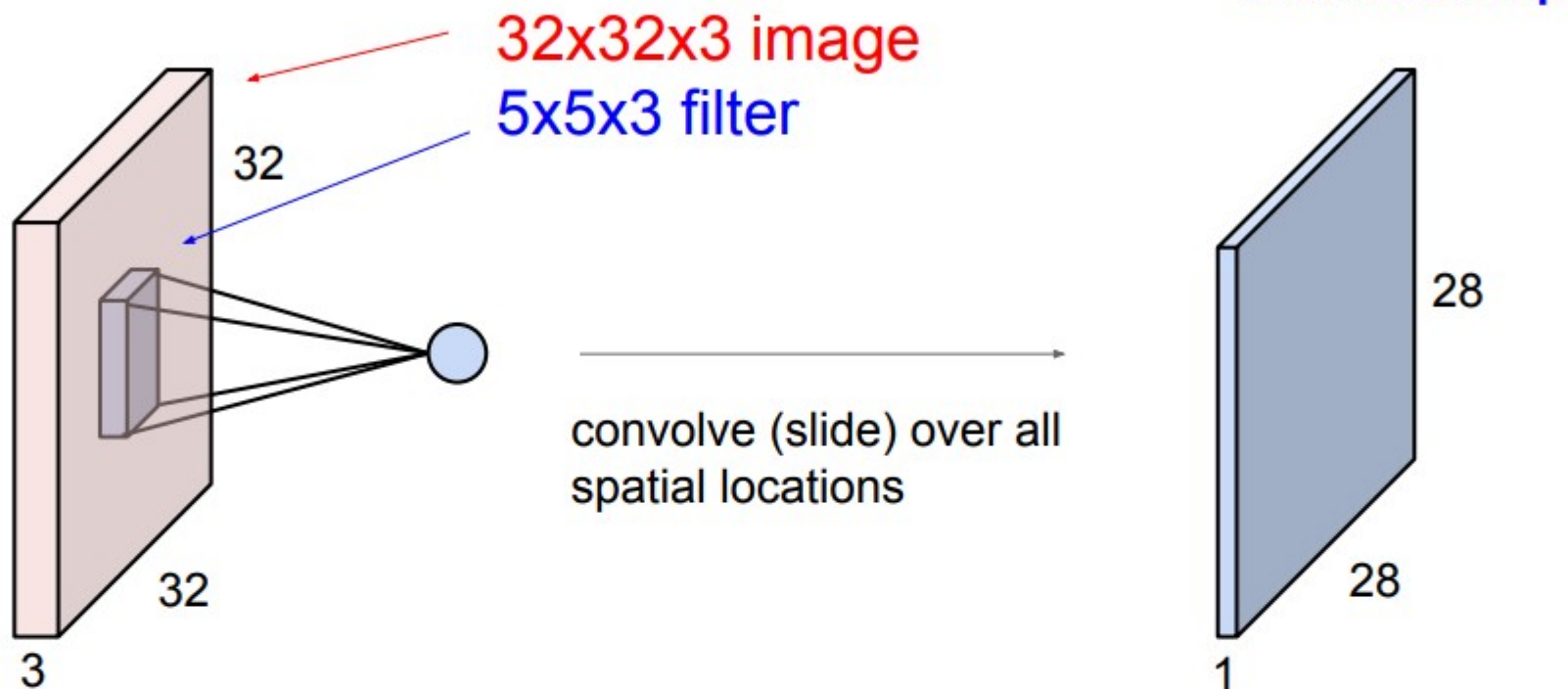


Convolve the filter with the image
i.e. “slide over the image spatially,
computing dot products”

Convolution Layer

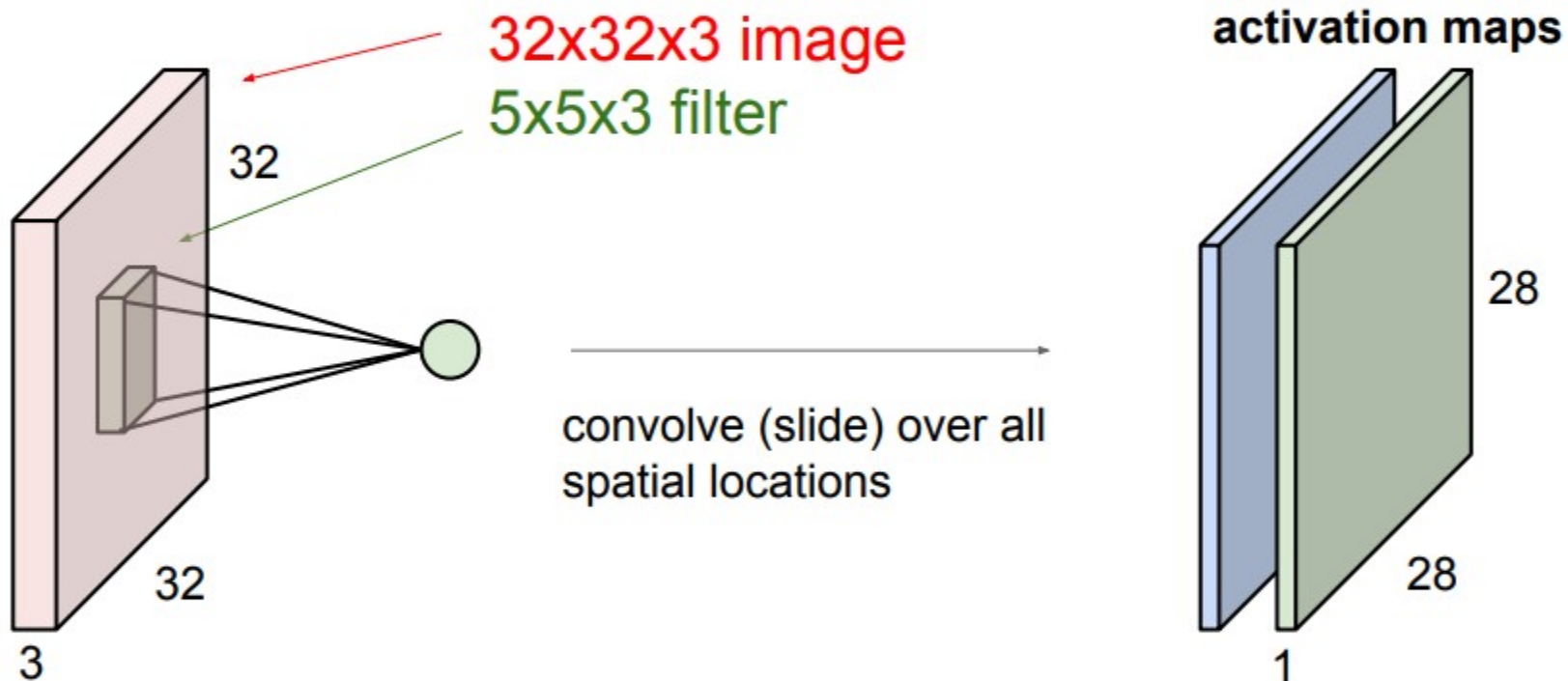


Convolution Layer



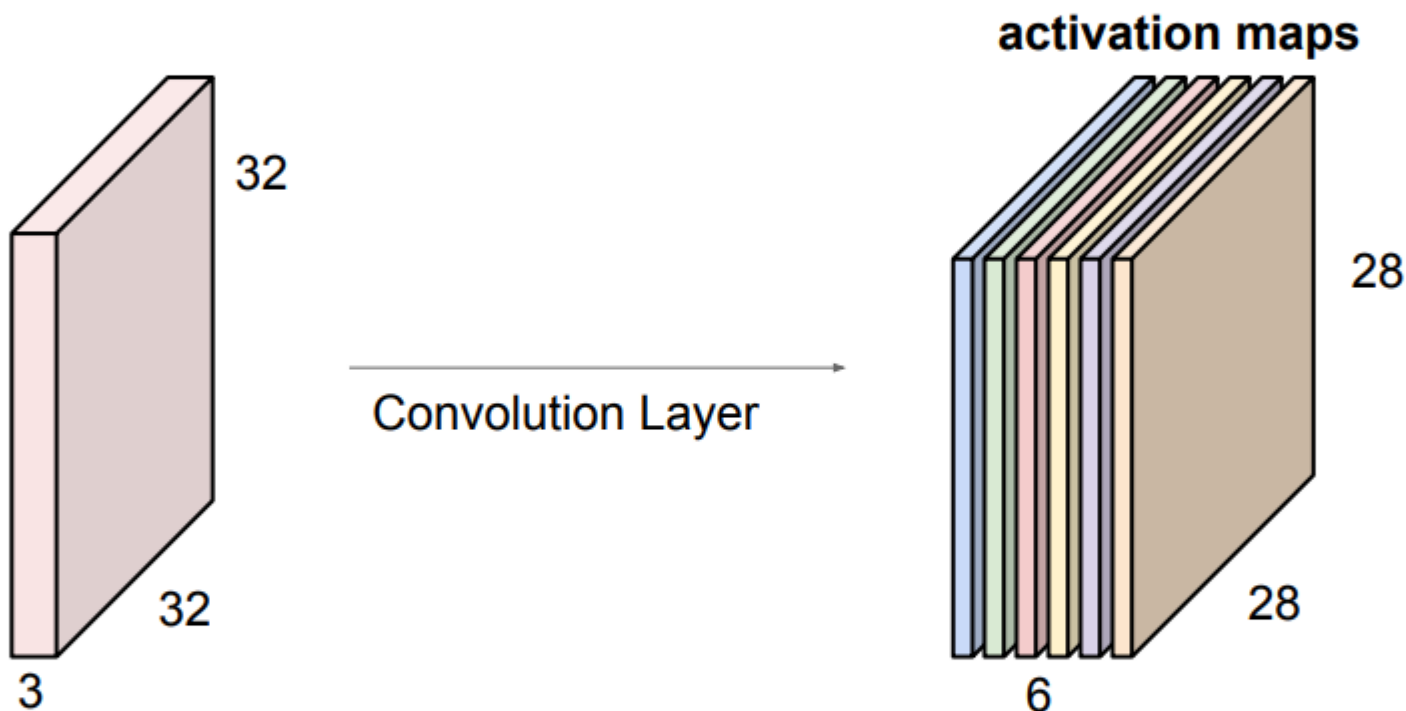
Convolution Layer

consider a second, **green** filter



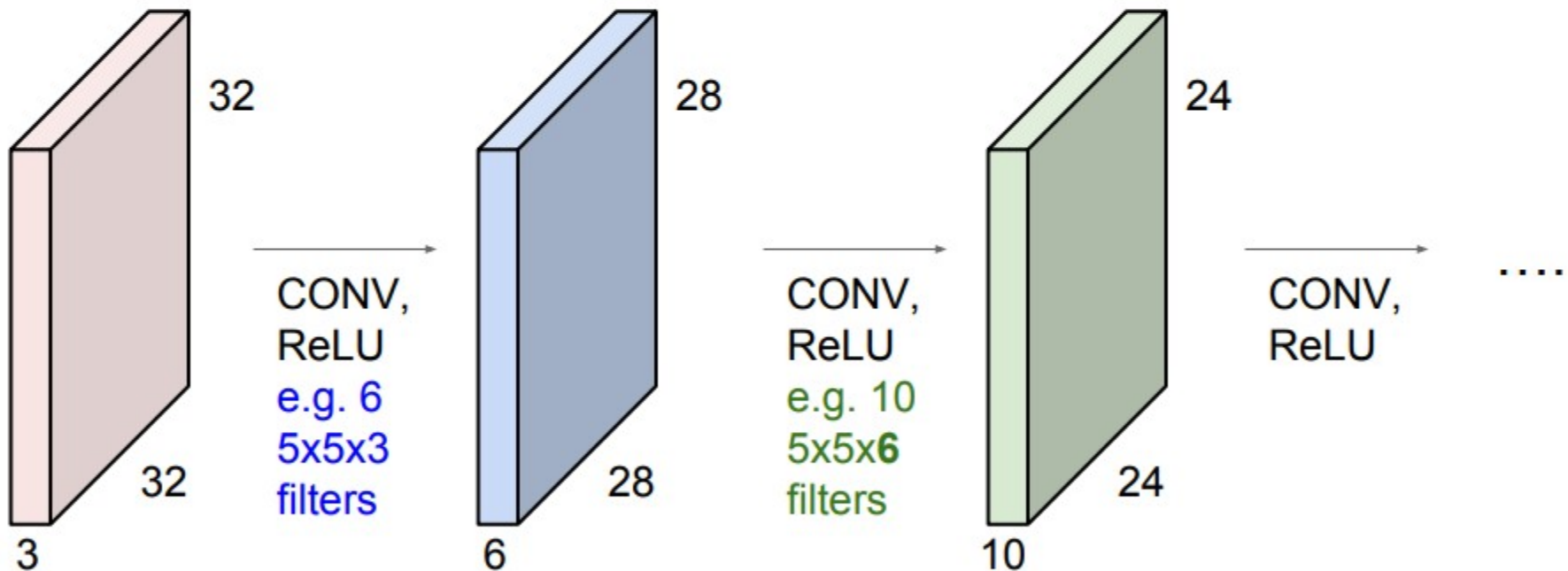
2D СВЕРТКА НАД 3D ТЕНЗОРАМИ

For example, if we had 6 5x5 filters, we'll get 6 separate activation maps:



We stack these up to get a “new image” of size 28x28x6!

Preview: ConvNet is a sequence of Convolutional Layers, interspersed with activation functions



Preview

[Zeiler and Fergus 2013]

Visualization of VGG-16 by Lane McIntosh. VGG-16 architecture from [Simonyan and Zisserman 2014].

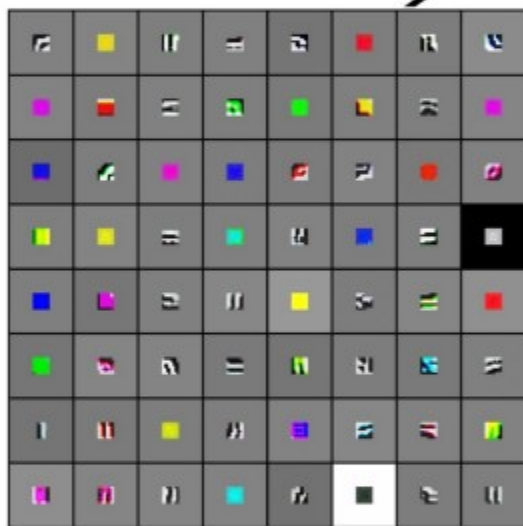


Low-level
features

Mid-level
features

High-level
features

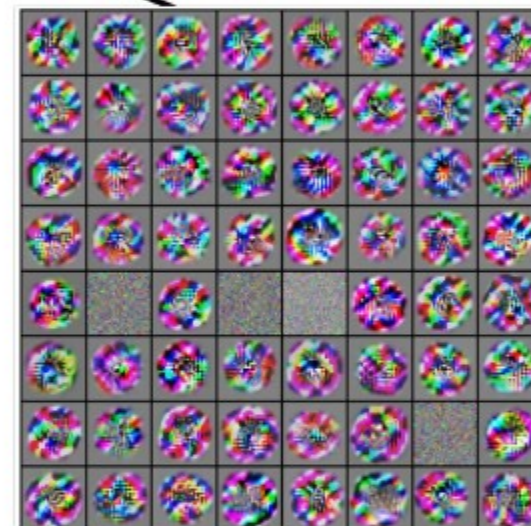
Linearly
separable
classifier



VGG-16 Conv1_1



VGG-16 Conv3_2



VGG-16 Conv5_3

In practice: Common to zero pad the border

0	0	0	0	0	0			
0								
0								
0								
0								

e.g. input 7x7

3x3 filter, applied with **stride 1**

pad with 1 pixel border => what is the output?

7x7 output!

in general, common to see CONV layers with stride 1, filters of size $F \times F$, and zero-padding with $(F-1)/2$. (will preserve size spatially)

e.g. $F = 3 \Rightarrow$ zero pad with 1

$F = 5 \Rightarrow$ zero pad with 2

$F = 7 \Rightarrow$ zero pad with 3

Common settings:

Summary. To summarize, the Conv Layer:

- Accepts a volume of size $W_1 \times H_1 \times D_1$
- Requires four hyperparameters:
 - Number of filters K ,
 - their spatial extent F ,
 - the stride S ,
 - the amount of zero padding P .
- Produces a volume of size $W_2 \times H_2 \times D_2$ where:
 - $W_2 = (W_1 - F + 2P)/S + 1$
 - $H_2 = (H_1 - F + 2P)/S + 1$ (i.e. width and height are computed equally by symmetry)
 - $D_2 = K$
- With parameter sharing, it introduces $F \cdot F \cdot D_1$ weights per filter, for a total of $(F \cdot F \cdot D_1) \cdot K$ weights and K biases.
- In the output volume, the d -th depth slice (of size $W_2 \times H_2$) is the result of performing a valid convolution of the d -th filter over the input volume with a stride of S , and then offset by d -th bias.

$K =$ (powers of 2, e.g. 32, 64, 128, 512)

- $F = 3, S = 1, P = 1$
- $F = 5, S = 1, P = 2$
- $F = 5, S = 2, P = ?$ (whatever fits)
- $F = 1, S = 1, P = 0$

MAX POOLING

Single depth slice

x ↑

1	1	2	4
5	6	7	8
3	2	1	0
1	2	3	4

→ y

max pool with 2x2 filters
and stride 2



6	8
3	4

Summary

- ConvNets stack CONV, POOL, FC layers
- Trend towards smaller filters and deeper architectures
- Trend towards getting rid of POOL/FC layers (just CONV)
- Typical architectures look like
 $[(\text{CONV-RELU}) * N - \text{POOL?}] * M - (\text{FC-RELU}) * K, \text{SOFTMAX}$
where N is usually up to ~5, M is large, $0 \leq K \leq 2$.
 - but recent advances such as ResNet/GoogLeNet challenge this paradigm