

Convolutional Neural Networks (CNNs)

Grigory Sapunov  CTO / Intento

Сверточная нейросеть

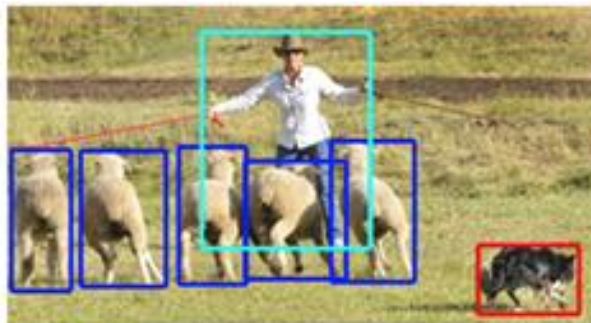
Свёрточная нейросеть (CNN) — это Feed-Forward сеть специального вида, для которой характерны:

- Weights sharing (даёт уменьшение числа параметров и возможность “учить” локальные признаки)
- Spatial pooling (даёт некоторую инвариантность к перемещениям)
- CNN работает с “объёмами” (тензорами)

Классические задачи для CNN



(a) classification



(b) detection



(c) segmentation

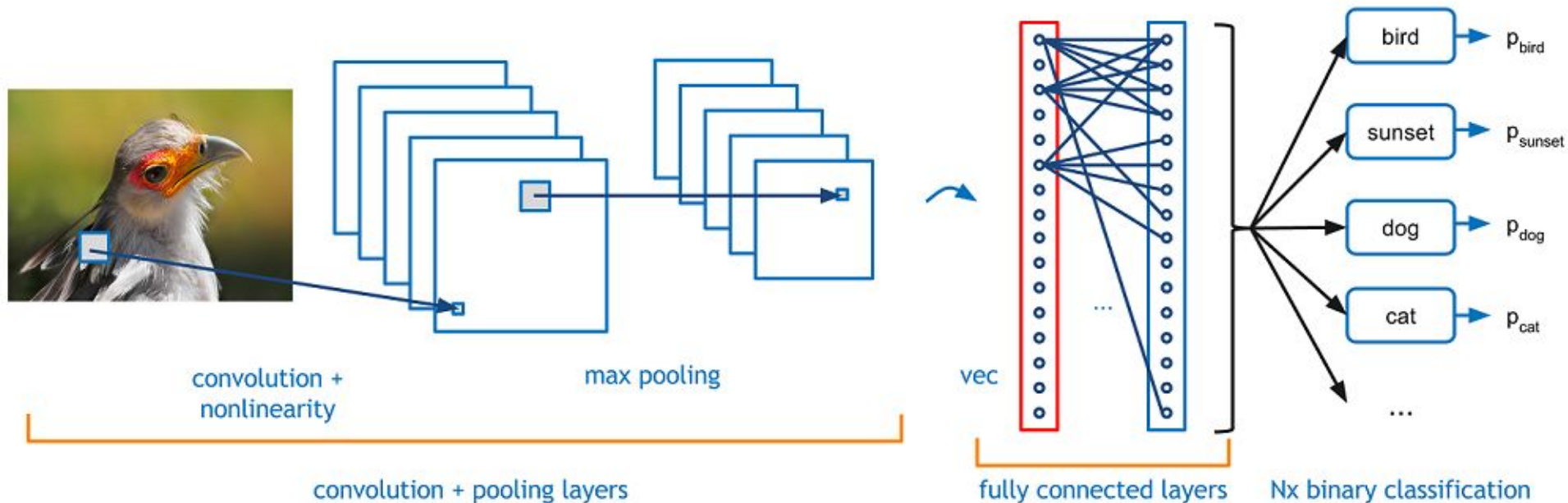
<https://research.facebook.com/blog/learning-to-segment/>

Другие задачи для CNN

- Трансформация изображений (перенос стиля, синтез изображений, ...)
- Обработка видео (spatio-temporal ...)
- Обработка пространственных данных
- Обработка звука (ASR, WaveNet, ...)
- Обработка текстов (классификация, перевод(!), ...)

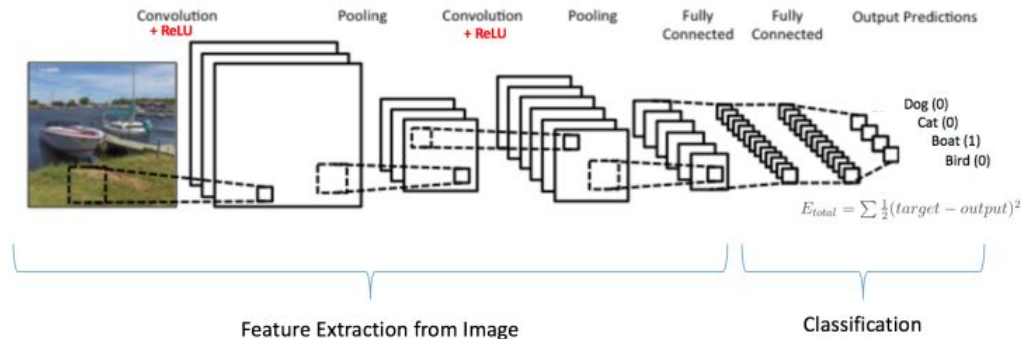
Свёрточная нейросеть: общий вид

Свёрточная нейросеть (CNN) — это Feed-Forward сеть специального вида:



Состав CNN

- **Свёрточные слои:** каждая плоскость в свёрточном слое — это один нейрон, реализующий операцию свёртки (convolution) и являющийся матричным фильтром небольшого размера (например, 5x5).
- **Слои субдискретизации** (subsampling, spatial pooling): уменьшают размер изображения (например, в 2 раза).
- **Полносвязные слои** (MLP) на выходе модели (используются для классификации).



Операция свёртки (convolution)

Свёртка — это математическая операция, комбинирующая два сигнала (функции от времени или пространства).

1D вариант:

$$y[n] = x[n] * h[n] = \sum_k x[k] * h[n - k] \text{ where } k \in [-\infty, +\infty]$$

$x[n]$ — сигнал

$h[n]$ — ядро (kernel) свёртки

Операция свёртки

2D вариант:

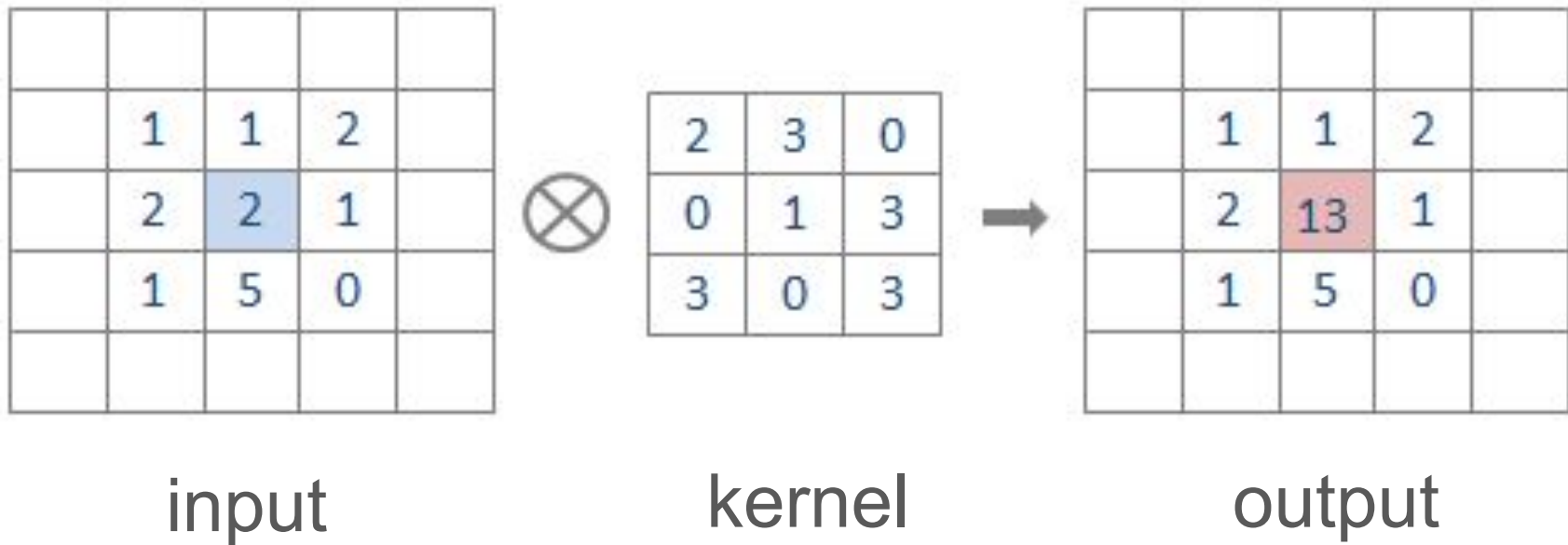
$$(f * g)(x, y) = \sum_{u=-\infty}^{+\infty} \sum_{v=-\infty}^{+\infty} f(u, v) \cdot g(x - u, y - v)$$

$f(x, y)$ — сигнал (например, изображение)

$g(u, v)$ — ядро свёртки

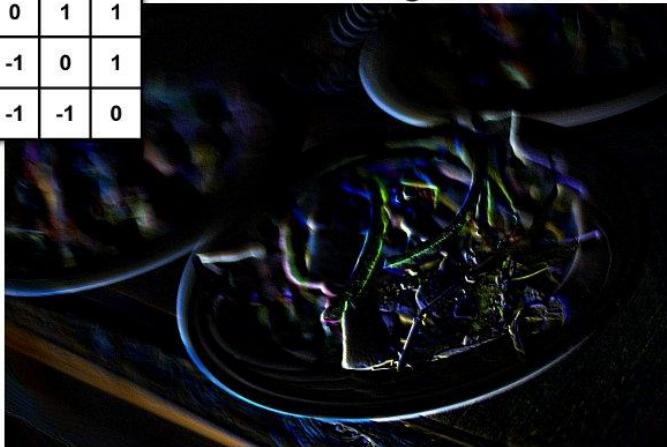
Визуализация операции свёртки

Знакомые по фотошопу фильтры blur, emboss, sharpen и другие — это именно матричные фильтры.

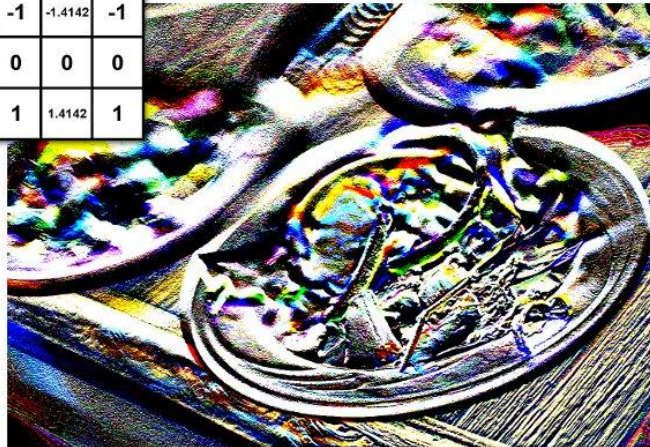


Diagonal Prewitt

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

**Horizontal Frei-Chen**

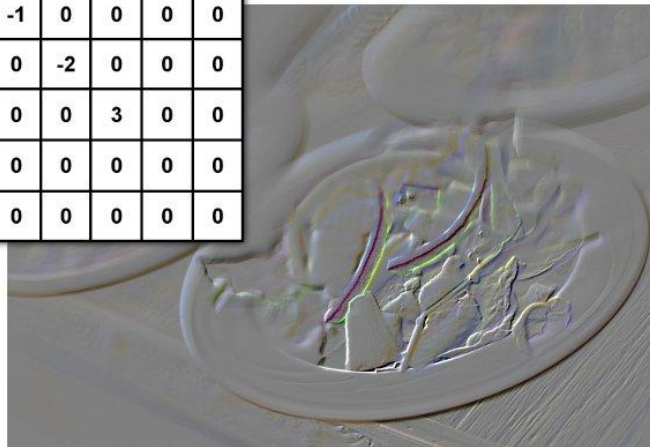
-1	-1.4142	-1
0	0	0
1	1.4142	1

**Sharpen**

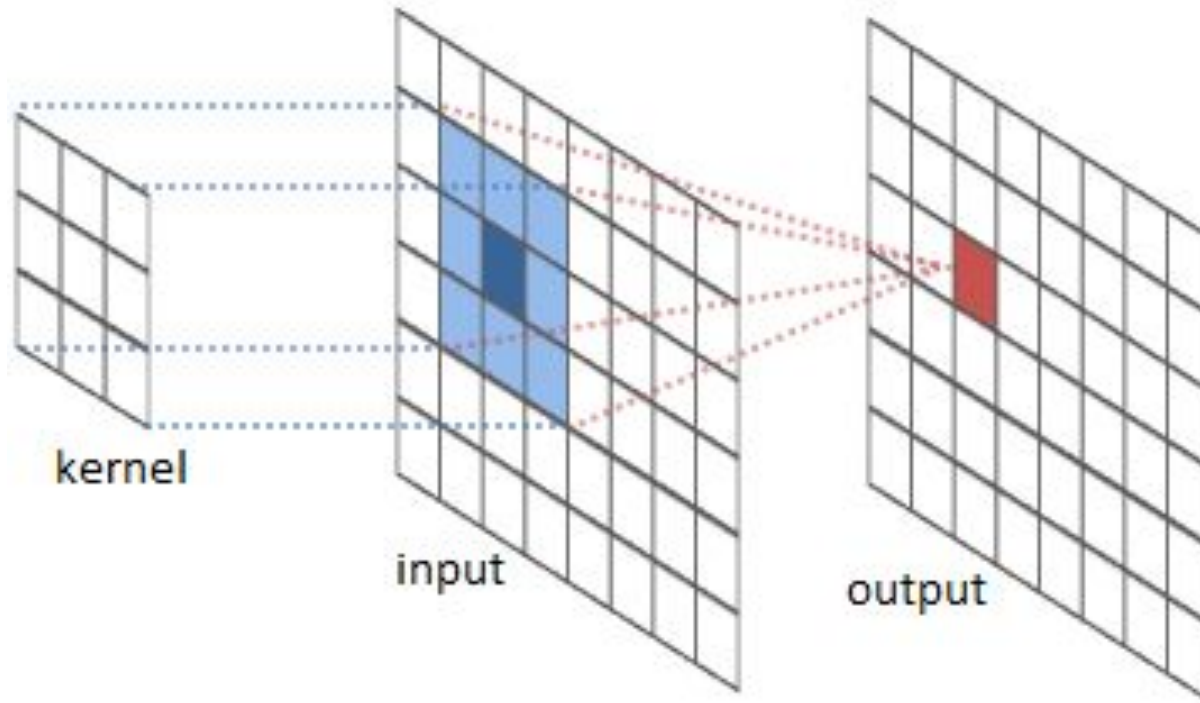
0	-1	0
-1	5	-1
0	-1	0

**Emboss**

-1	0	0	0	0
0	-2	0	0	0
0	0	3	0	0
0	0	0	0	0
0	0	0	0	0



Визуализация операции свёртки



Визуализация двумерной операции свёртки

1 _{x1}	1 _{x0}	1 _{x1}	0	0
0 _{x0}	1 _{x1}	1 _{x0}	1	0
0 _{x1}	0 _{x0}	1 _{x1}	1	1
0	0	1	1	0
0	1	1	0	0

Image

4		

Convolved
Feature

Важные параметры для свёртки

- Размер фильтра (размер ядра)
 - 3x3
 - 5x5
 - ...

3X3

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

3X3

0	-1	0
-1	4	-1
0	-1	0

5X5

0	0	1	0	0
0	1	2	1	0
1	2	-16	2	1
0	1	2	1	0
0	0	1	0	0

7X7

0	0	1	1	1	0	0
0	1	3	3	3	1	0
1	3	0	-7	0	3	1
1	3	-7	-24	-7	3	1
1	3	0	-7	0	3	1
0	1	3	3	3	1	0
0	0	1	1	1	0	0

9X9

0	0	3	2	2	2	3	0	0
0	2	3	5	5	5	3	2	0
3	3	5	3	0	3	5	3	3
2	5	3	-12	-23	-12	3	5	2
2	5	0	-23	-40	-23	0	5	2
2	5	3	-12	-23	-12	3	5	2
3	3	5	3	0	3	5	3	3
0	2	3	5	5	5	3	2	0
0	0	3	2	2	2	3	0	0

Важные параметры для свёртки

- **Padding** (заполнение на границах изображения)

Если не заполнять, то размер обработанного изображения будет меньше исходного.

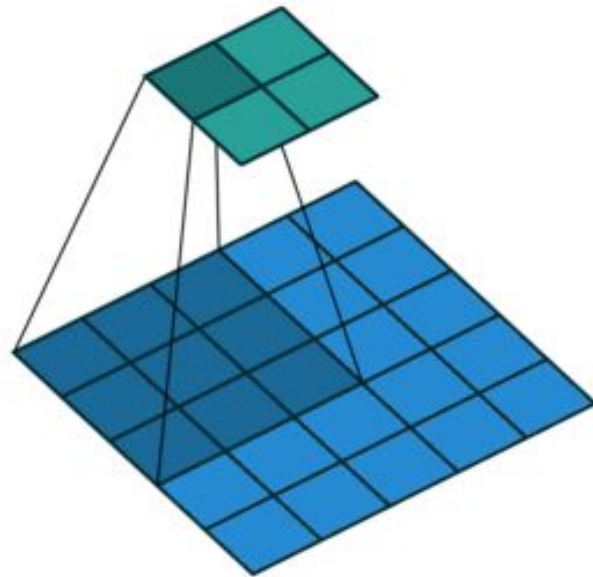
- Пример с padding = 1

0 ₂	0 ₀	0 ₁	0	0	0	0
0 ₁	2 ₀	2 ₀	3	3	3	0
0 ₀	0 ₁	1 ₁	3	0	3	0
0	2	3	0	1	3	0
0	3	3	2	1	2	0
0	3	3	0	2	3	0
0	0	0	0	0	0	0

1	6	5
7	10	9
7	10	8

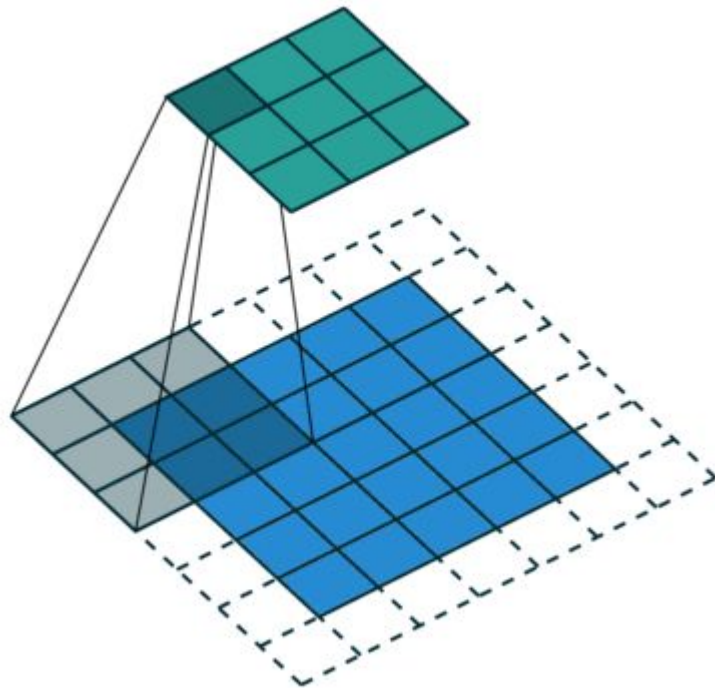
Важные параметры для свёртки

- **Stride** (шаг при движении фильтра)
 - Пример с $\text{stride} = 2$

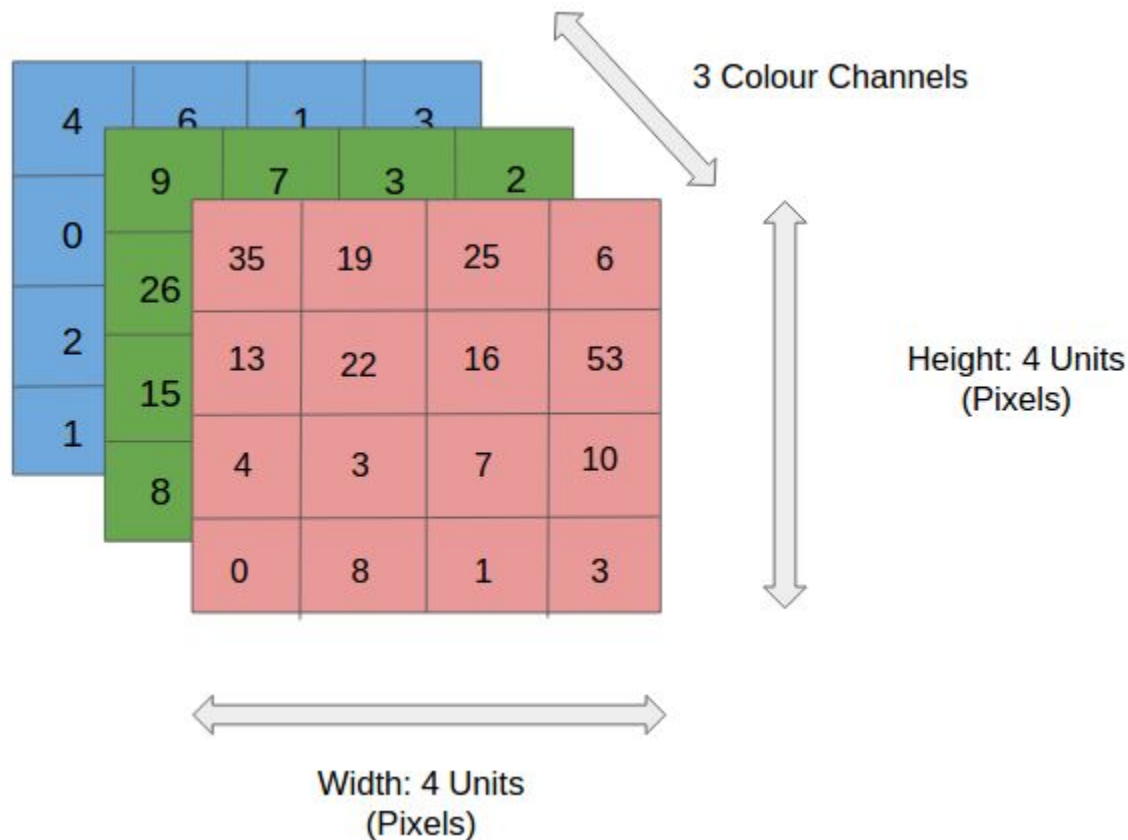


Важные параметры для свёртки

- Пример
 - image size = 5
 - kernel size = 3
 - padding = 1
 - stride = 2

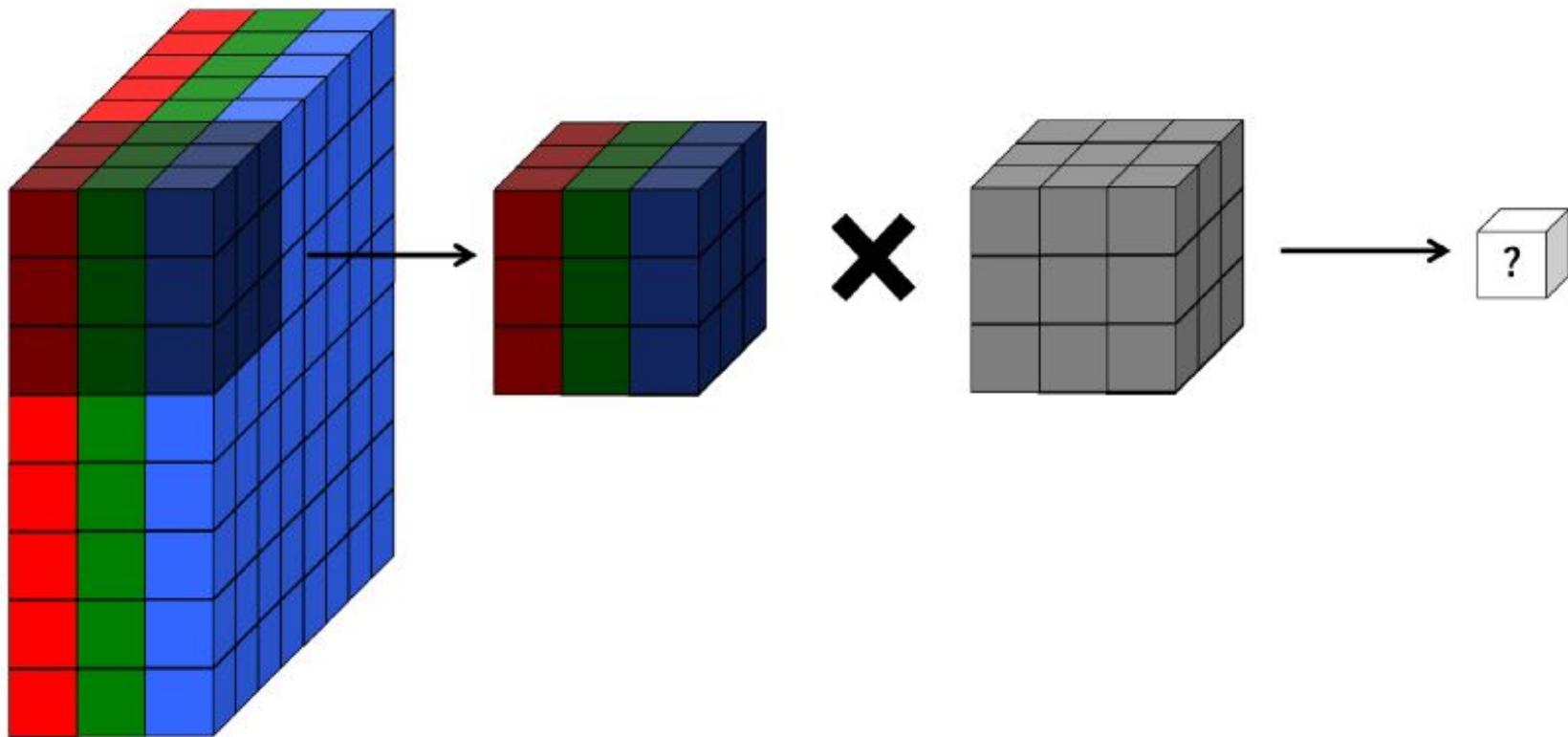


Входные данные — 3-х мерный сигнал



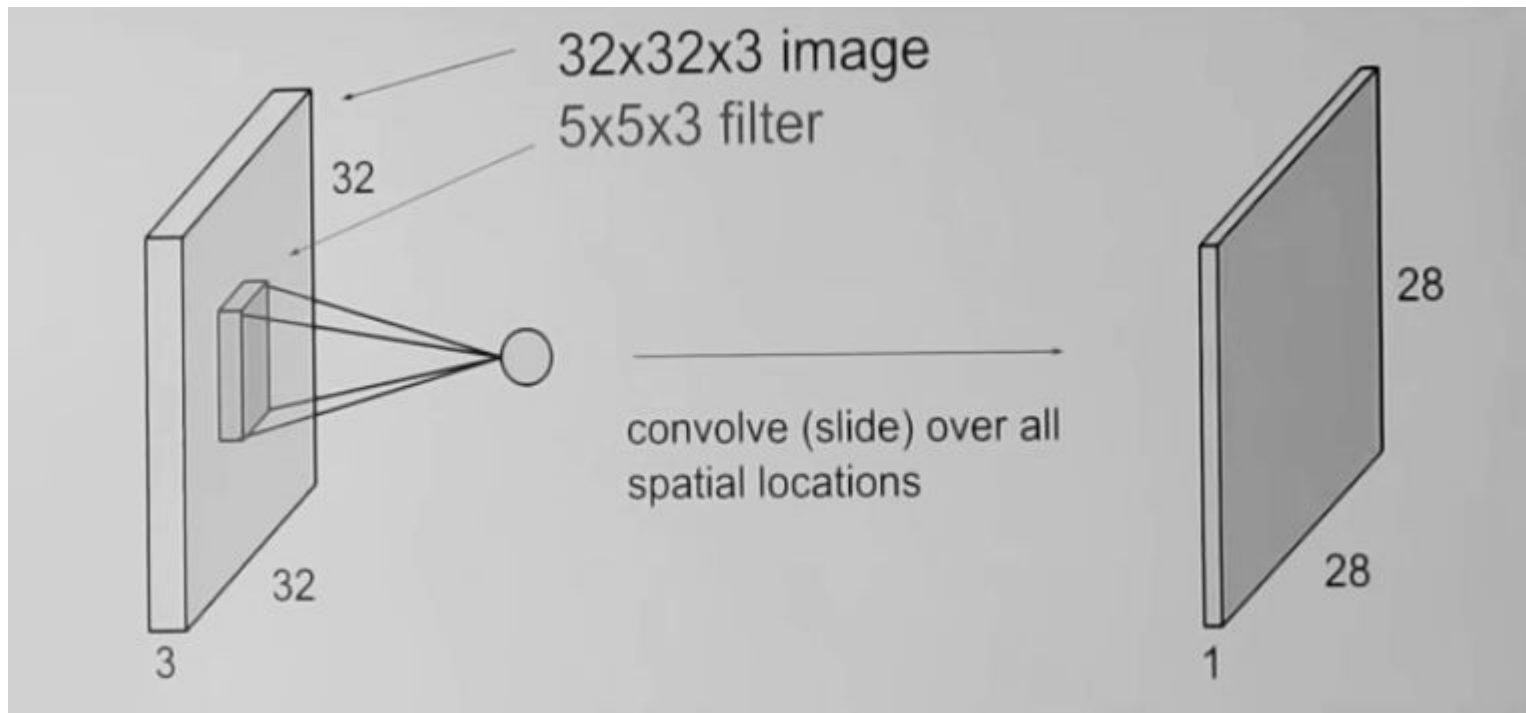
Свёртка работает над объёмами

В реальности фильтр трёхмерный.



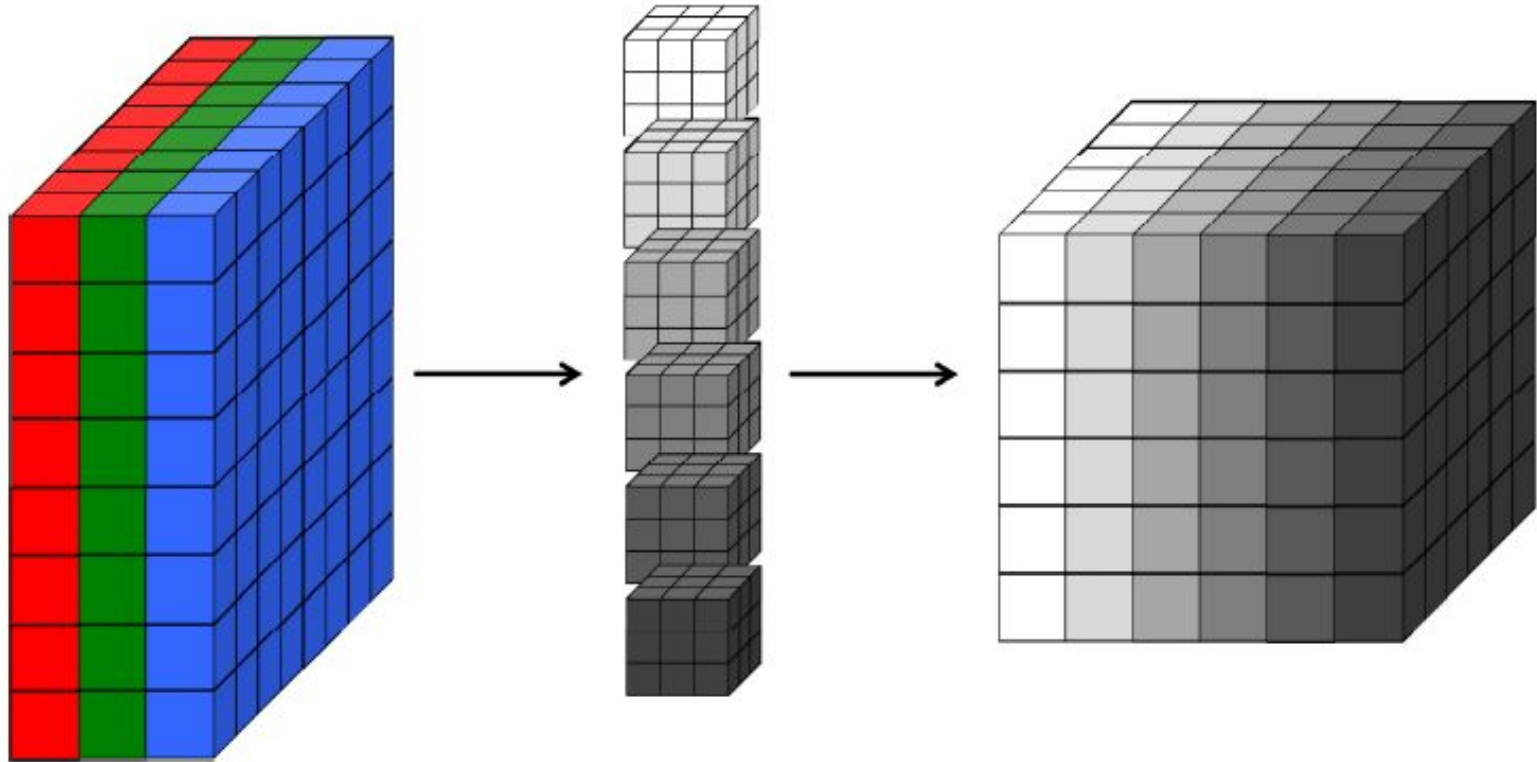
Свёртка работает над объёмами

Один фильтр генерирует одну “плоскость” из изображения.



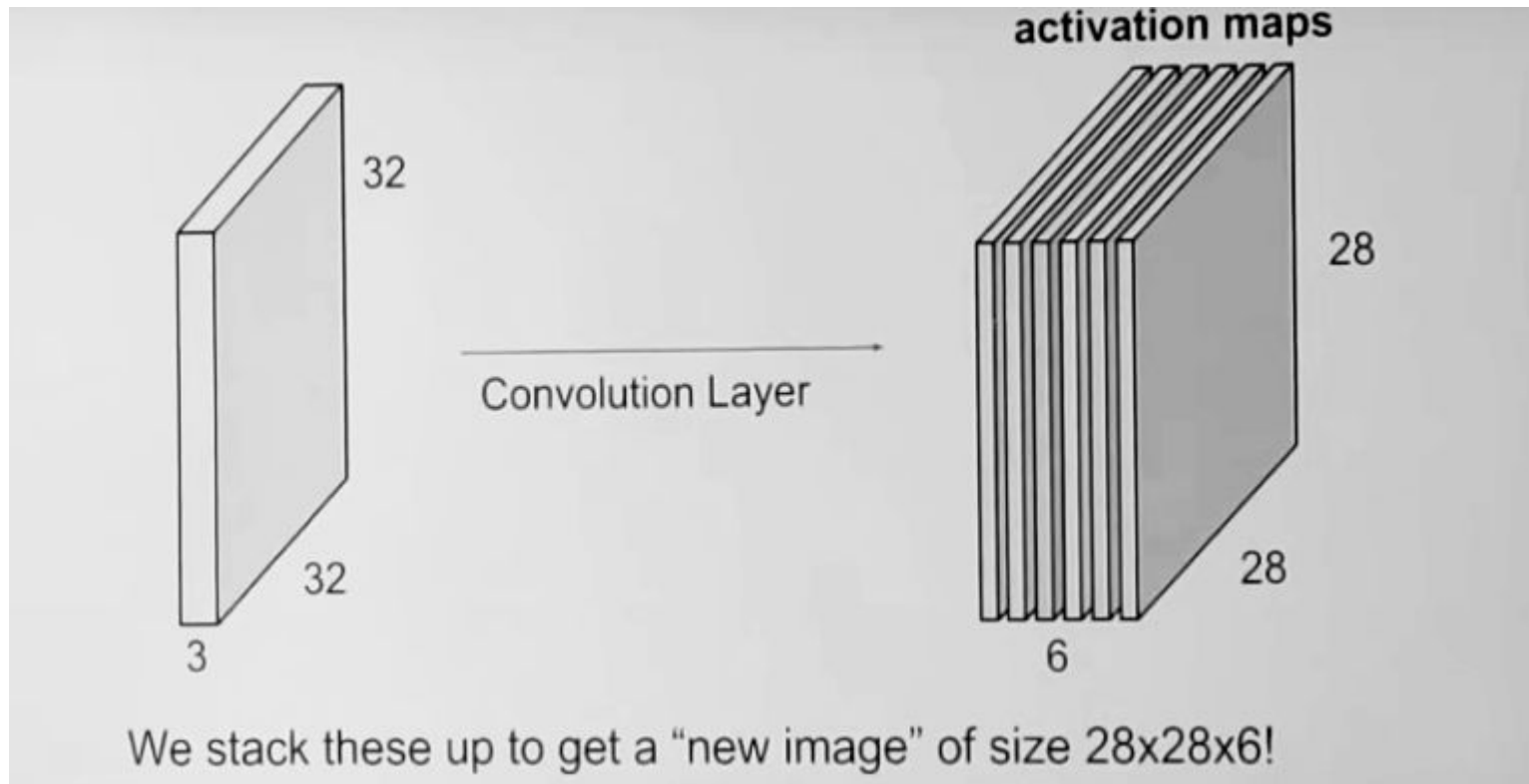
Свёртка работает с объёмами

Несколько фильтров (образующих один слой) создают “объём”

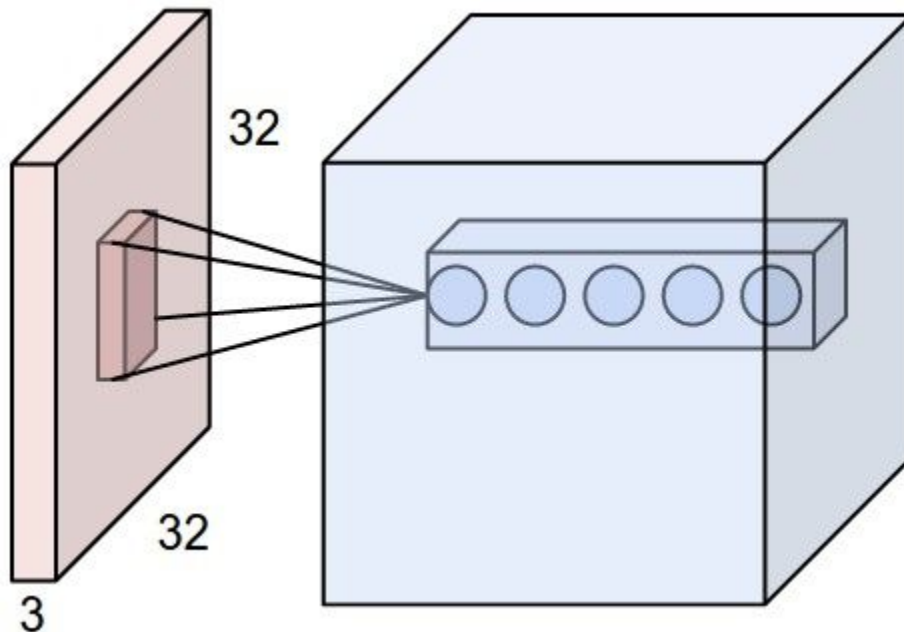


Свёртка работает с объёмами

Несколько фильтров (образующих один слой) создают “объём”

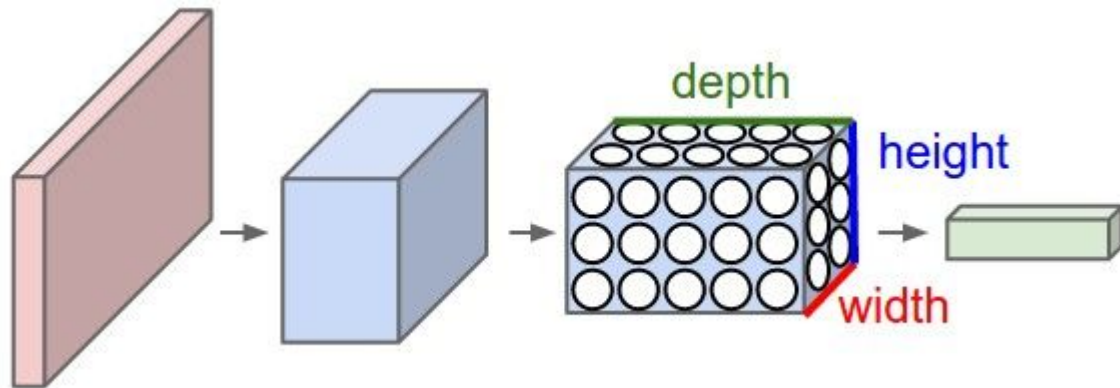
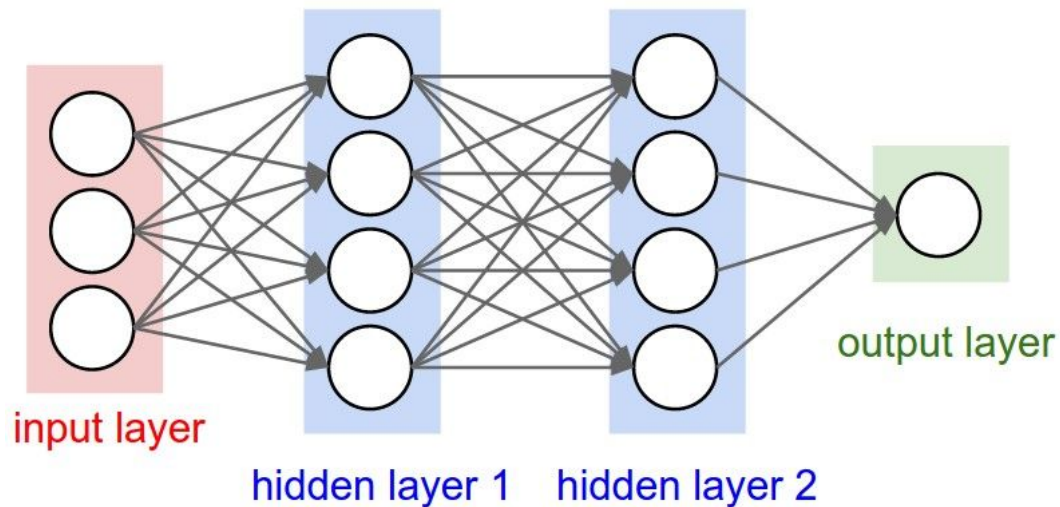


CNN: Свёрточный слой (5 нейронов)

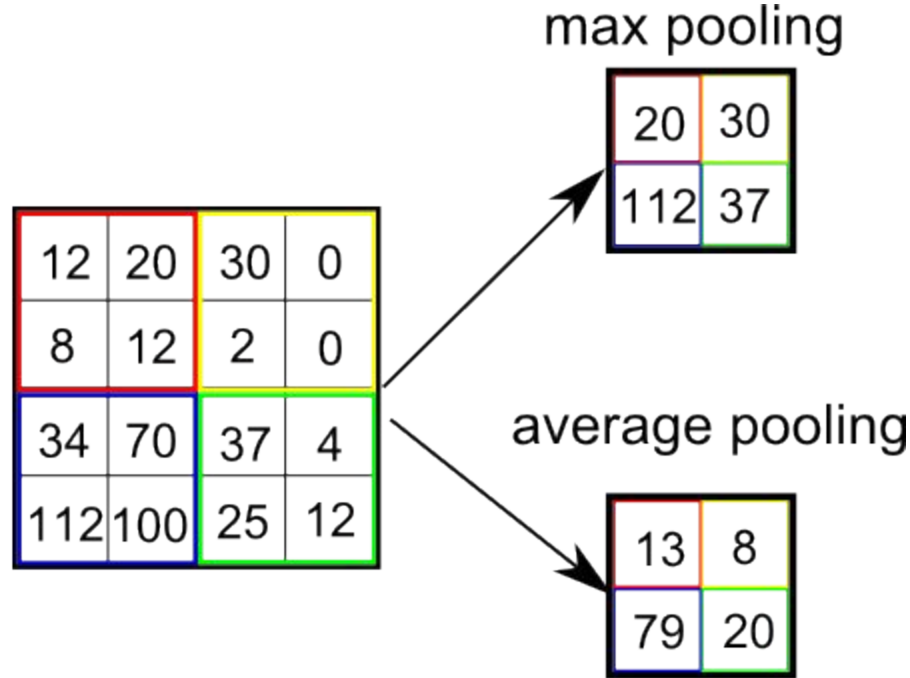


Веса нейронов — это коэффициенты ядра свёртки. Каждая “обучаемая” свёртка выделяет одинаковые локальные признаки во всех частях изображения.

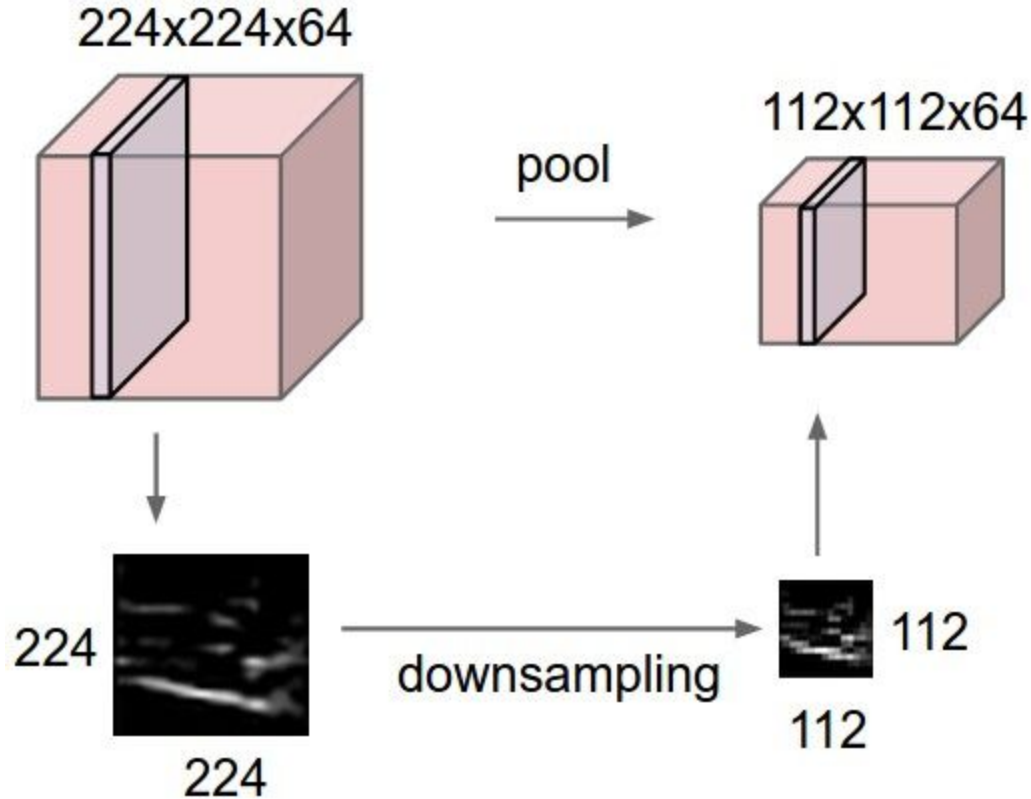
Свёртка работает над объёмами



Операция pooling (max pool, avg pool)

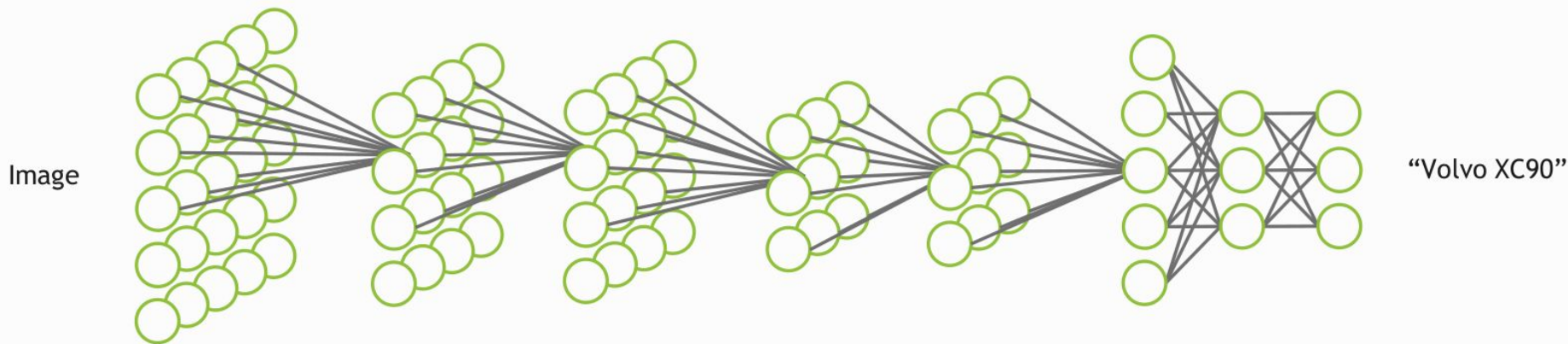
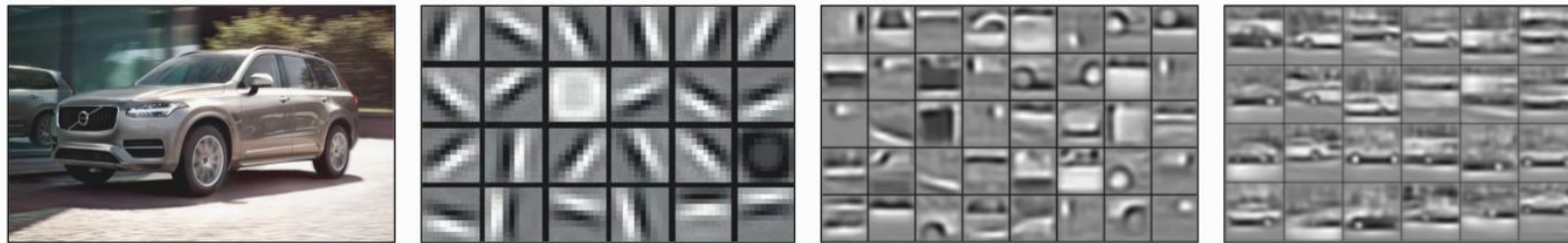


CNN: Pooling слой (downsampling)



Свёрточная нейросеть

Свёрточные слои учат иерархические признаки для изображений, а spatial pooling даёт некоторую инвариантность к перемещениям.



У CNN меньше параметров, чем у FNN

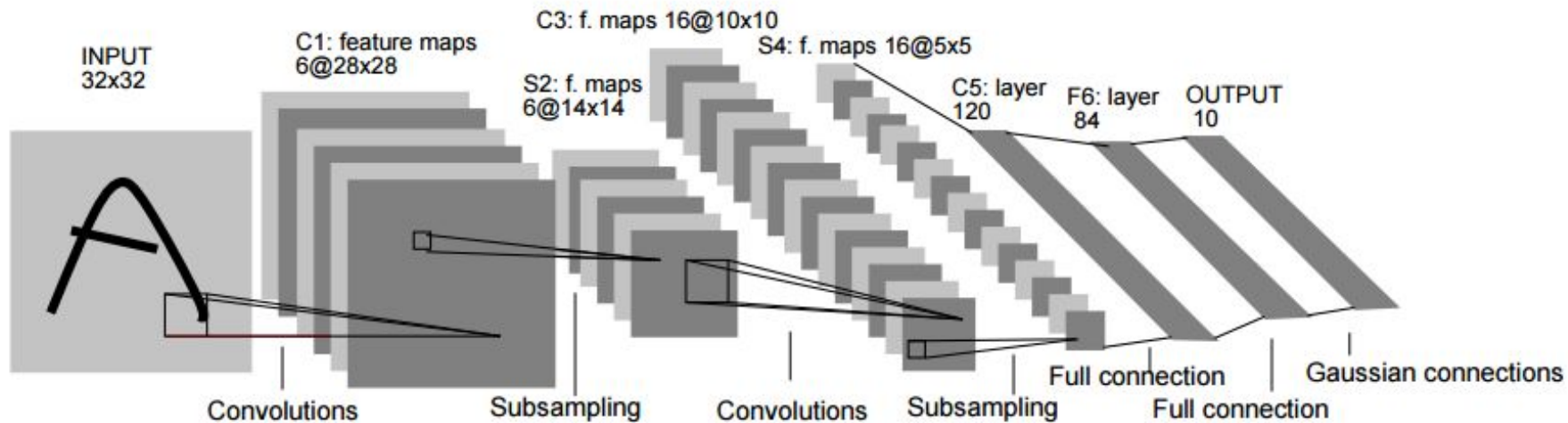
CNN

- вход ч/б картинка 100x100
- три свёрточных слоя по 100 плоскостей каждый (conv 5x5 и subsampling 2)
- выход: 10 классов
- число параметров примерно **650K** ($5*5*1*100 + 5*5*100*100 + 5*5*100*100 + 12*12*100*10$)

FNN

- вход: ч/б картинка 100x100
- три скрытых слоя по 100 нейронов каждый
- выход: 10 классов
- число параметров примерно **1M** ($10000*100 + 100*100 + 100*100 + 100*10$)

LeNet-5



Jupyter notebook example