

Лекция 1. Свёрточные нейронные сети

Александр Юрьевич Авдюшенко

МКН СПбГУ

17 февраля 2022



Hubel & Wiesel (1959)

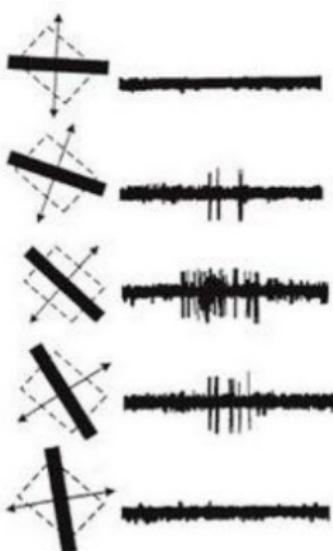
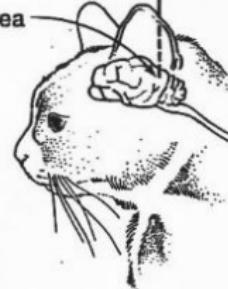
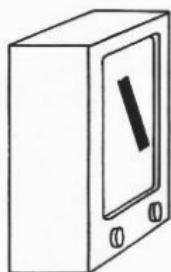
История

The Nobel Prize in Physiology or Medicine, 1981

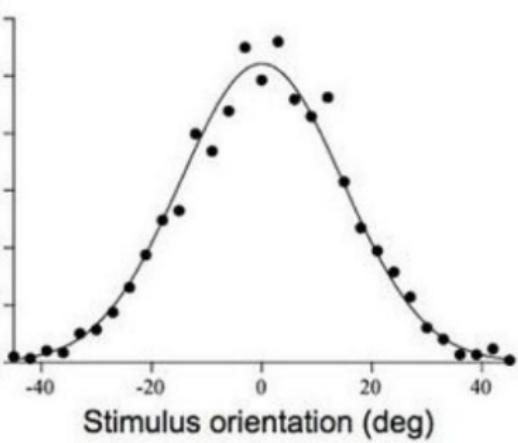
Electrical signal
from brain

Recording electrode →

Visual area
of brain



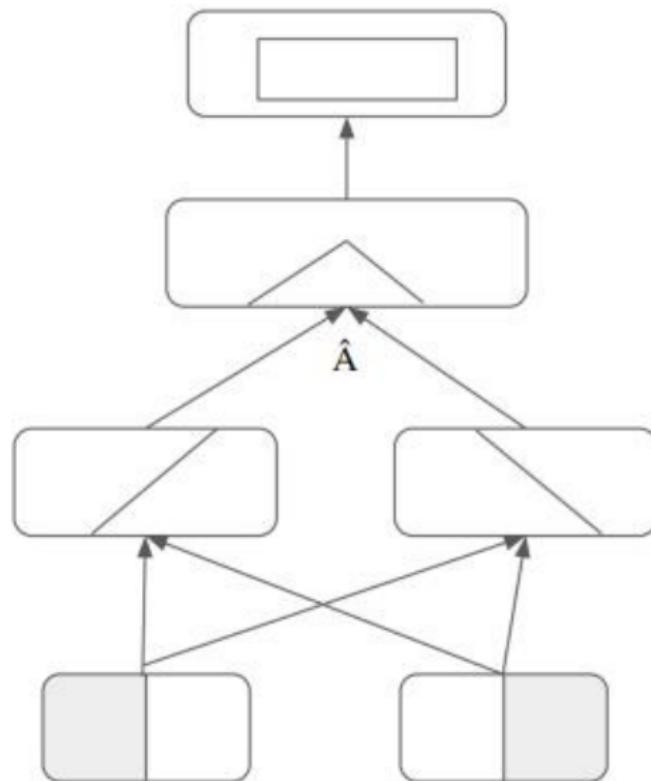
Neural response (spikes/sec)



Stimulus orientation (deg)

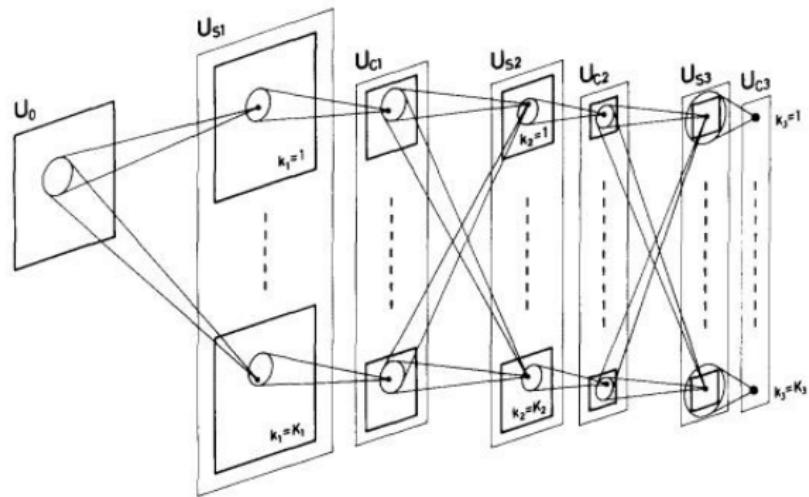
Идея иерархической организации

История



Fukushima (1980)

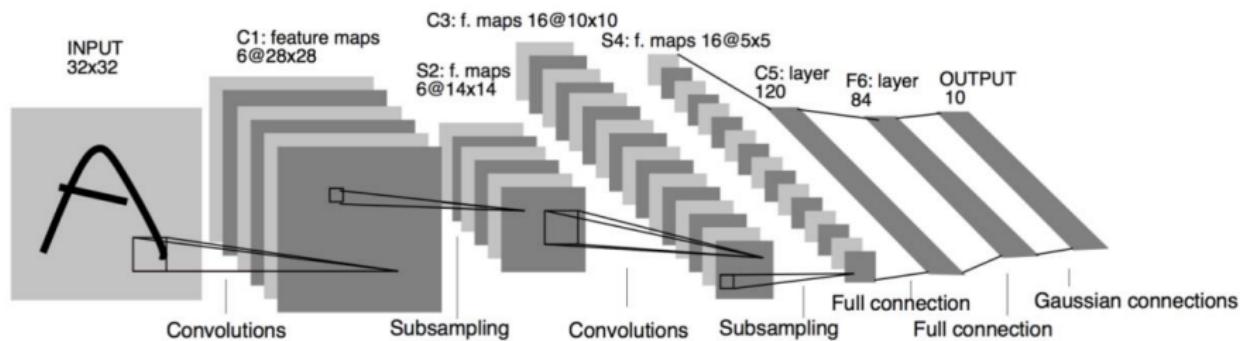
История



Уже использовались «свёртки» и активации, но без градиентного спуска и обучения с учителем

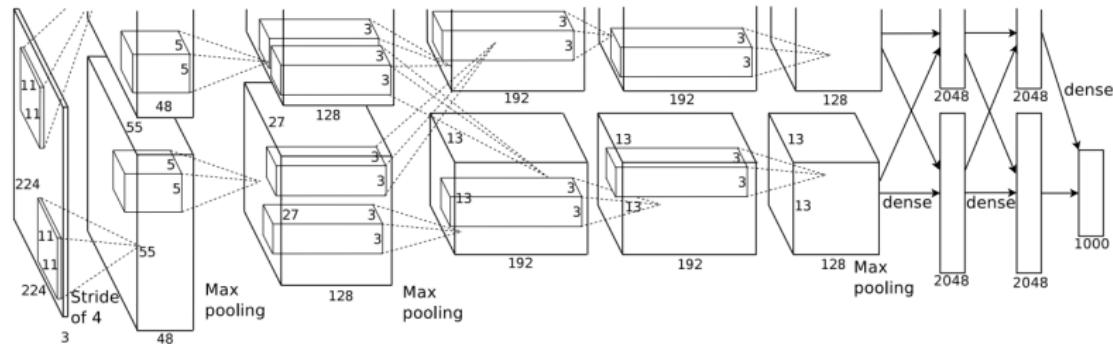
Lekun, Bottou, Bengio, Haffner (1998)

Первый успех



Победитель конкурса ImageNet того времени

AlexNet CNN Architecture



Линейная модель

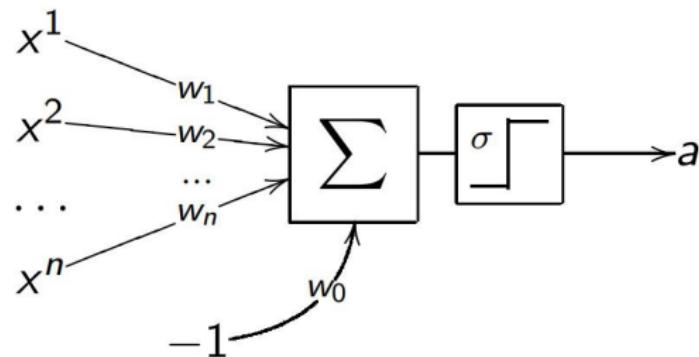
Напоминание

$f_j : X \rightarrow \mathbb{R}$ — числовые признаки

$$a(x, w) = \sigma(\langle w, x \rangle) = \sigma \left(\sum_{j=1}^n w_j f_j(x) - w_0 \right),$$

где $w_0, w_1, \dots, w_n \in \mathbb{R}$ — веса признаков

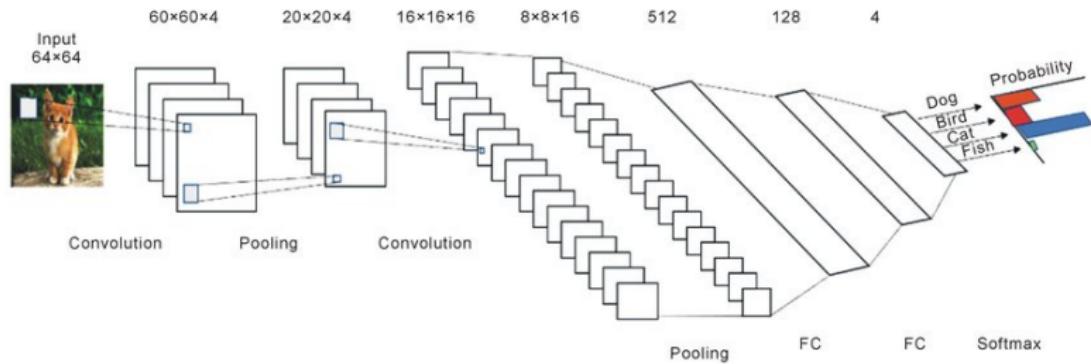
$\sigma(z)$ — функция активации, например, $\text{sign}(z)$, $\frac{1}{1+e^{-z}}$, $(z)_+$



Нейронная сеть как комбинация линейных моделей



Свёрточная нейронная сеть



Операция свёртки (convolution) в нейронных сетях

Свёртка в нейросетях — сумма произведений элементов

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |



Input Image

| | | |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 0 | 0 |
| 0 | 1 | 1 |

Feature
Detector



| | | | |
|---|--|--|--|
| 0 | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Feature Map

Вопрос

Почему «свёртка»?

Замечание

При реализации свёртки эффективно умножается матрица на вектор. Вот, например, [статья с реализацией Winograd transformation в cuDNN](#).

Пример операции свёртки

Ядро $3 \times 3 \times 3$ (Weight \times Height \times Channel numbers)

Дополнение (padding) и шаг (stride)

Расширение (dilation=2)

Считаем размер выхода

- ▶ Filter size = 3x3 -> 3
- ▶ Input size = 5x5 -> 5
- ▶ Stride = 1x1 -> 1
- ▶ Padding = 0x0 -> 0

$$\text{Output size} = (I - F + 2*P)/S + 1 = (5 - 3 + 2*0)/1 + 1 = 3$$

Output size = 3 -> 3x3

Пулинг (pooling, субдискретизация)

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

max pool with 2x2
window and stride 2

| | |
|---|---|
| 6 | 8 |
| 3 | 4 |

Сигмоидальные функции активации

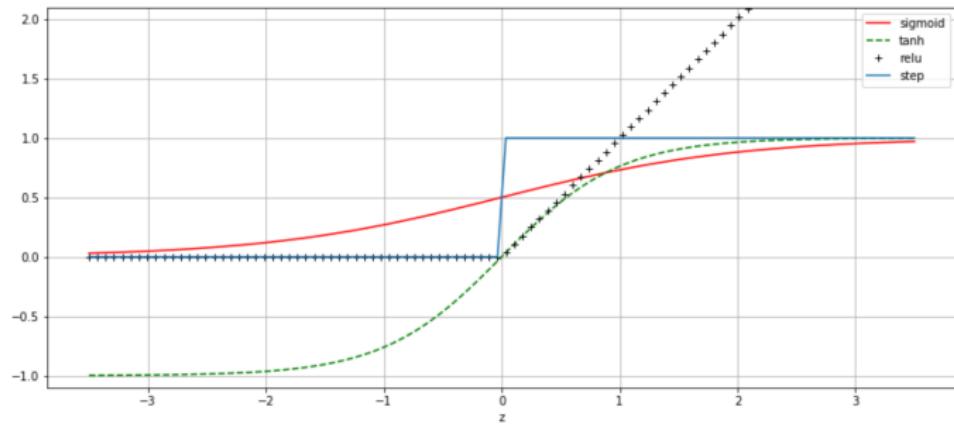
Функции активации

- ▶ Логистический сигмоид $\sigma(z) = \frac{1}{1+\exp(-z)}$
- ▶ Гиперболический тангенс $\tanh(z) = \frac{\exp(z)-\exp(-z)}{\exp(z)+\exp(-z)}$
- ▶ непрерывные аппроксимации пороговых функций
- ▶ могут приводить к затуханию градиентов или «параличу» сети

Вопрос

Какие недостатки у логистического сигмоида?

Посмотрим на графики



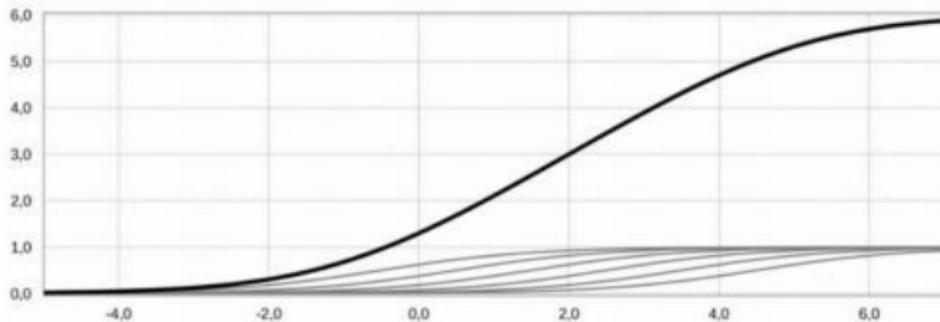
ReLU — Rectified Linear Unit

$$\text{ReLU}(z) = \max(0, z)$$

Мотивация:

$$\sigma(5) \approx 0.9933, \sigma(10) \approx 0.9999$$

$$f(x) = \sigma\left(x + \frac{1}{2}\right) + \sigma\left(x - \frac{1}{2}\right) + \sigma\left(x - \frac{3}{2}\right) + \sigma\left(x - \frac{5}{2}\right) + \dots$$



Николенко С., Кадурин А., Архангельская Е. «Глубокое обучение», Питер, 2018. (стр. 107-113)

ReLU — Rectified Linear Unit

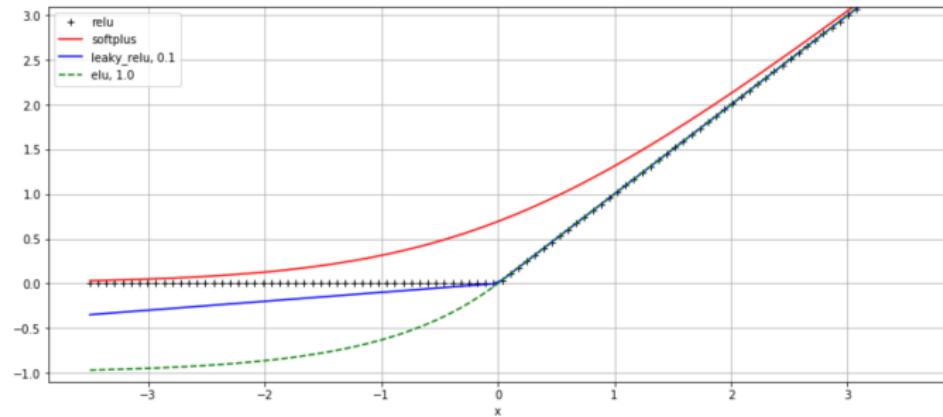
$$\int \sigma(x)dx = \log(1 + e^x) + C$$

Получается, $f(x)$ — это риманова сумма вот такого интеграла:

$$\int_{1/2}^{\infty} \sigma\left(x + \frac{1}{2} - y\right)dy$$

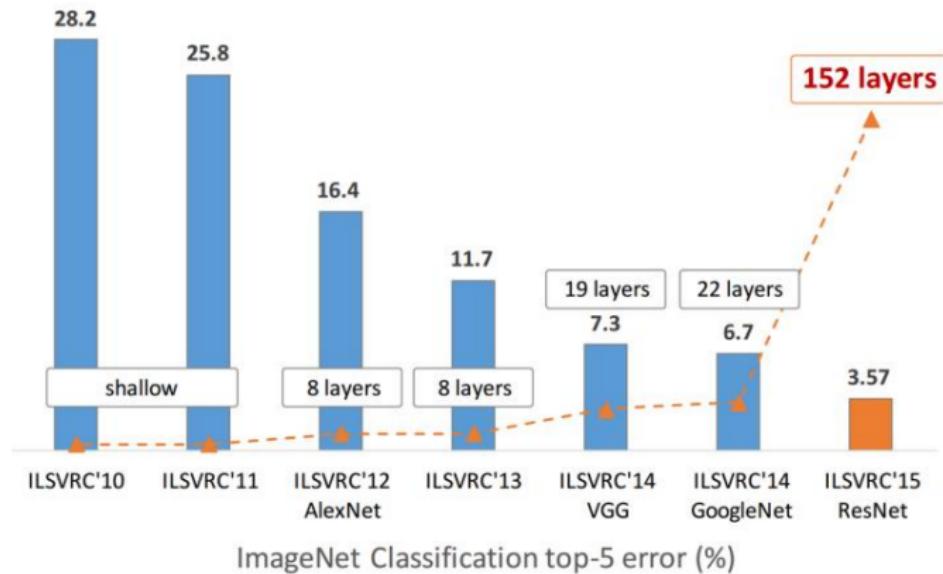
$$\begin{aligned} f(x) &= \sum_{i=0}^{\infty} \sigma\left(x + \frac{1}{2} - i\right) \approx \int_{1/2}^{\infty} \sigma\left(x + \frac{1}{2} - y\right)dy = \\ &= [-\log(1+\exp(x+\frac{1}{2}-y))]_{y=1/2}^{y=\infty} = \log(1+\exp(x)) = \text{Softplus}(x) \end{aligned}$$

Посмотрим на графики



Развитие свёрточных сетей

или краткая история ImageNet

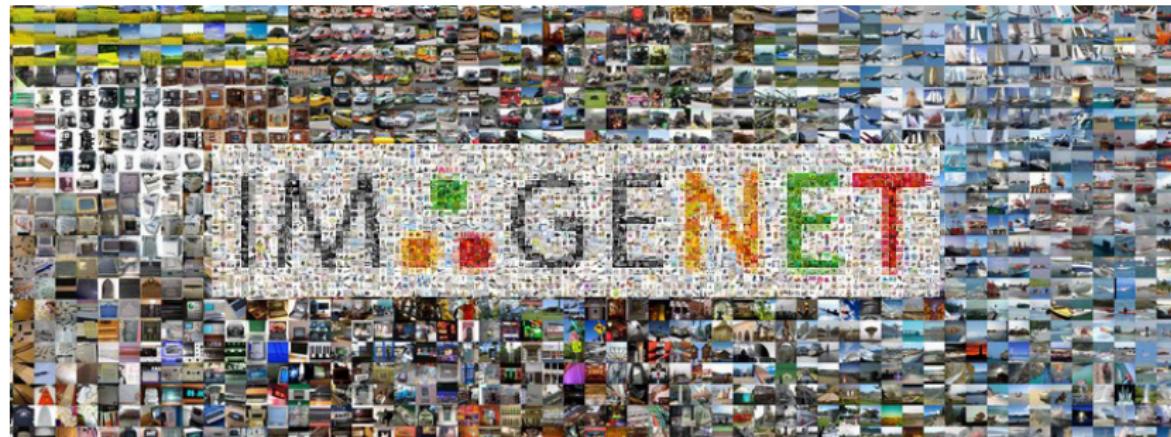


Вопрос

Что такое ImageNet?

Проект массивной базы данных аннотированных изображений:

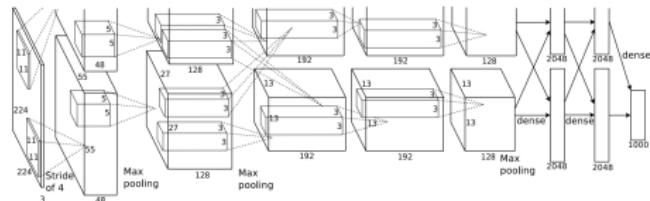
- ▶ 14M+ изображений
- ▶ 20K+ классов



AlexNet (Krizhevsky, Sutskever, Hinton, 2012)

Победитель конкурса ImageNet того времени

AlexNet CNN Architecture



- ▶ ReLU
- ▶ аугментация данных
- ▶ *dropout 0.5*
- ▶ *batch normalization* (размер батча 128)
- ▶ SGD Momentum 0.9
- ▶ L2 регуляризация 5e-4
- ▶ Learning rate 1e-2, потом уменьшение в 10 раз после стабилизации качества на тесте

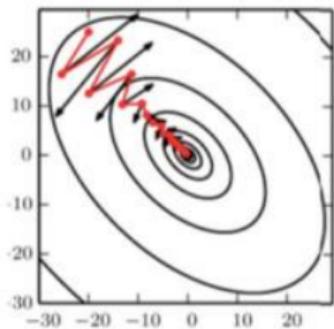
Финальное качество top5 на ImageNet — 25.8% -> 16.4%

Метод накопления импульса (momentum method)

Метод накопления импульса [Б.Т.Поляк, 1964] — экспоненциальное скользящее среднее градиента по $\frac{1}{1-\gamma}$ последним итерациям:

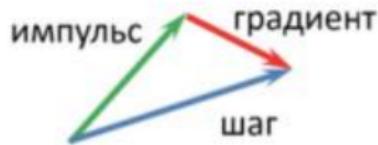
$$\nu = \gamma \nu + (1 - \gamma) \mathcal{L}'_i(w)$$

$$w = w - \eta \nu$$



Поправка Нестерова (1983)

$$\begin{aligned}\nu &= \gamma\nu + (1 - \gamma)\mathcal{L}'_i(w - \eta\gamma\nu) \\ w &= w - \eta\nu\end{aligned}$$



Резюме

- ▶ Свёрточные сети очень хорошо подходят для обработки изображений
- ▶ По-видимому немного похожи на механизмы биологического зрения
- ▶ При этом гибкие и вычислительно эффективные
- ▶ Сегодня фактический стандарт для задач компьютерного зрения (классификация, детекция, сегментация, генерация)
- ▶ Будет на следующей лекции
 - ▶ различные алгоритмы оптимизации: adam, RMSProp
 - ▶ dropout
 - ▶ выбор начального приближения
 - ▶ ResNet и WideResNet

Что ещё можно посмотреть?

- ▶ Демо от Андрея Карпаты
- ▶ Есть курс Стенфорда «CS231n: сверточные нейронные сети для визуального распознавания»: <http://cs231n.github.io/>