

СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

Сергей Николенко

СПбГУ — Санкт-Петербург

11 сентября 2026 г.

Random facts:



- 11 сентября 1943 г. на стекольном заводе в Гусь-Хрустальном был выпущен первый советский гранёный стакан; форму, по легенде, придумала Вера Мухина
- 11 сентября 1962 г. The Beatles закончили запись первого сингла «P.S. I Love You»; на барабанах был сессионный музыкант Энди Уайт, а Ринго Старру дали маракасы
- 11 сентября 1973 г. в результате переворота в Чили погиб президент Альенде; вместе с ним закончился и «Киберсин» — проект централизованного управления экономикой, который строил там британский кибернетик Стаффорд Бир
- 11 сентября — дата невесёлая: ураган на Ньюфаундленде в 1847 г., землетрясение в Крыму в 1927 г., пожары в Пакистане в 2012 г. и, конечно, 2001 год
- 11 сентября — Национальный день Каталонии, учреждённый в память о взятии Барселоны во время войны за испанское наследство в 1713 г.
- 11 сентября — важная дата для флота Петра I: в 1714 г. в этот день произошло Гангутское сражение, а в 1720 г. — сражение при Гренгаме

КРАТКОЕ РЕЗЮМЕ ПРЕДЫДУЩИХ СЕРИЙ

- В прошлый раз мы разобрали три сюжета, без которых глубокие сети не обучаются: регуляризацию и дропаут, инициализацию весов и нормализацию по мини-батчам.
- Всё это — про полносвязные сети: каждый нейрон смотрит на все входы сразу.
- Для картинок это плохо: соседние пиксели связаны сильнее далёких, а сдвиг объекта на пару пикселей не должен менять ответ.
- Сегодня — архитектура, которая это учитывает: *свёрточные сети*.
- Начнём с того, как это подсмотрели у зрительной коры, а закончим первой сетью, которая выиграла ImageNet и с которой началась вся революция, — AlexNet.

СВЁРТОЧНЫЕ СЕТИ: ИДЕЯ

ЗРИТЕЛЬНАЯ КОРА

- Зрительная кора давно известна.
- Но до середины XX века изучать могли только на очень «высоком» уровне: болезни, травмы, агнозии.
- Хьюбел и Визель: исследования отдельных нейронов, на что они активируются.



ЗРИТЕЛЬНАЯ КОРА

- Оказалось, что нейроны «ближних» к сетчатке уровней активируются на простые формы.
- Вообще оказалось, что зрительная кора делится на уровни: V1, V2, V3...
- Но от V3 и выше Хьюбел и Визель уже мудро не стали исследовать.

A Experimental setup

Light bar stimulus
projected on screen

Recording from visual cortex

Record



B

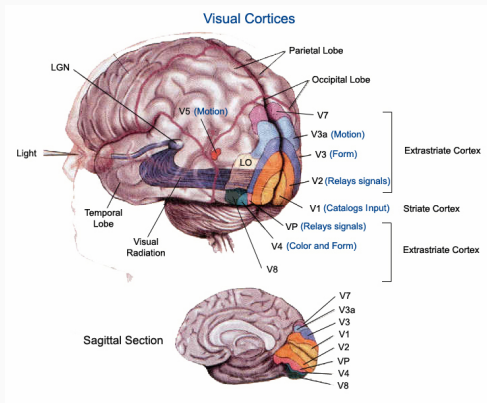
Stimulus
orientation

Stimulus
presented

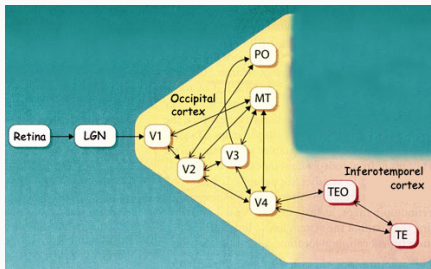


ЗРИТЕЛЬНАЯ КОРА

- Сетчатка → латеральное коленчатое тело (LGN, lateral geniculate nucleus) → зрительная кора.
- И там есть чёткая структура, организованная в уровни от V1 до V4-V6.

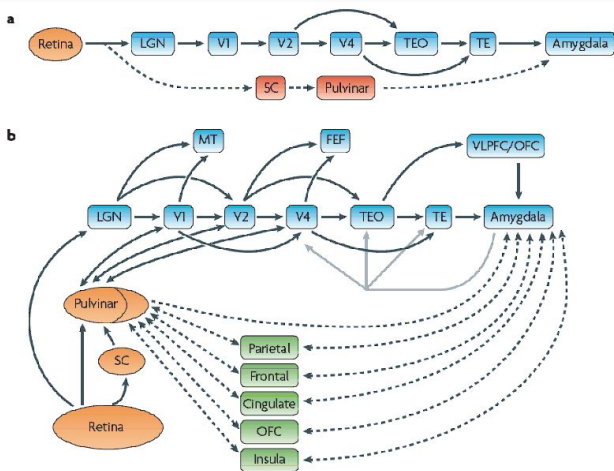


- Ну, вроде того. Мозг – сложная штука.



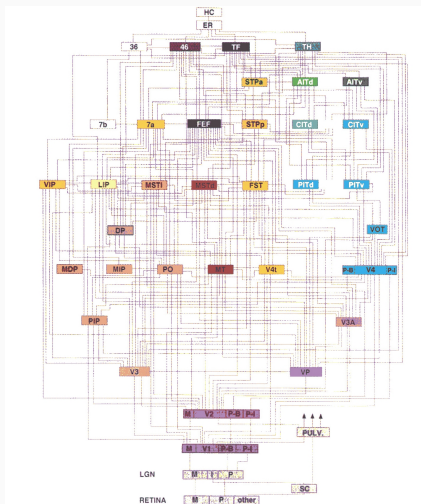
ЗРИТЕЛЬНАЯ КОРА

- Правда сложная.



ЗРИТЕЛЬНАЯ КОРА

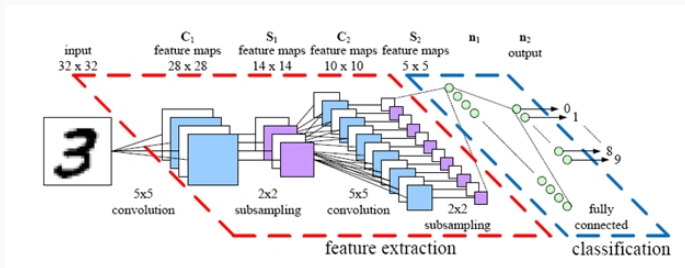
- Честно-честно.



- Но некие условные уровни там действительно есть:
 - V1 – локальные признаки, топографические карты;
 - V2 – больше локальных признаков, бинокулярное зрение;
 - V3 – цвет, текстура, первые результаты сегментации;
 - V4 – геометрические формы, силуэты; это самый важный уровень для *внимания*;
 - V5 – распознаёт движения объектов, сегментированных на V4;
 - V6 – обобщает данные со всей картинки, wide-field stimulation; например, обрабатывает наши собственные движения;
 - V7 (наверное) – распознаёт сложные объекты, например человеческие лица.
- What pathway: V2 – V4, распознавание форм и объектов.
- Where pathway: V2 – V5 – V6, распознавание движений, управление глазами и руками.

СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- Эти идеи нашли отражение в искусственных сетях.
- Свёртки впервые появились в *Neocognitron* (Fukushima, 1979; 1980).
- В современной форме – в группе ЛеКуна во второй половине 1980-х.
- Главная идея: хочется применить одну и ту же операцию к разным частям изображения.



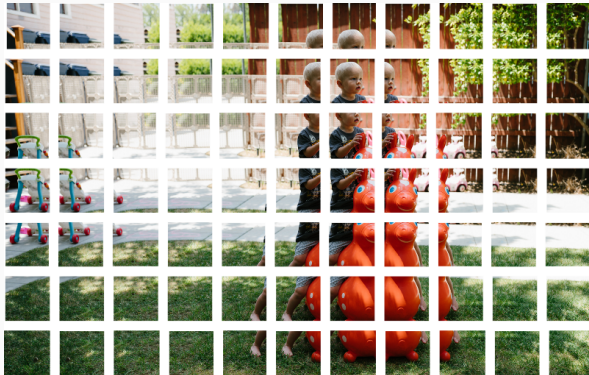
СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- Разобьём изображение на окна:



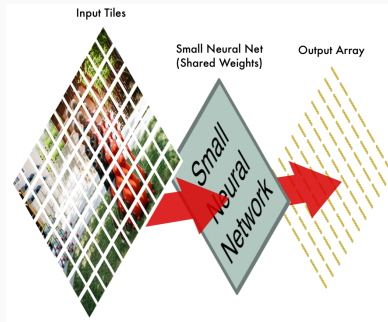
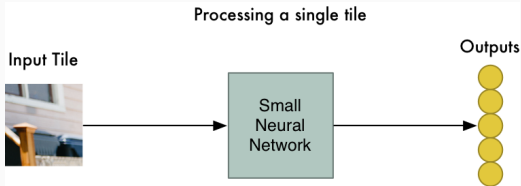
СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- Разобьём изображение на окна:



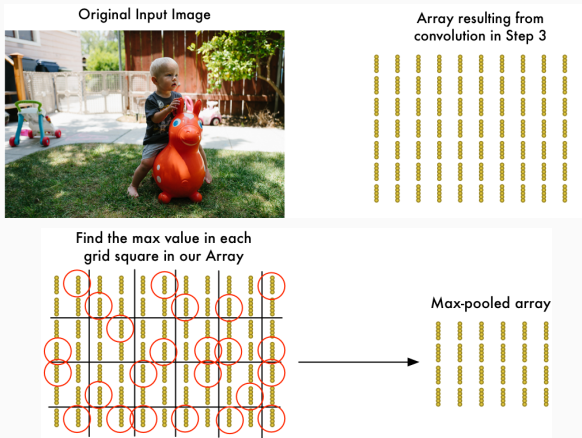
СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- Применим маленькую нейронную сеть к каждому окну:



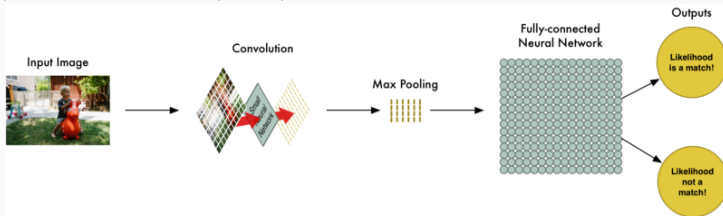
СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- Затем сожмём результат, сделав субдискретизацию (pooling; например, max-pooling) по маленьким окнам:

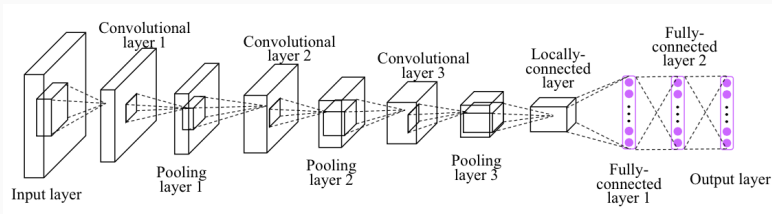


СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- А затем уже используем эти признаки, чтобы делать предсказания, например, полносвязной сетью:

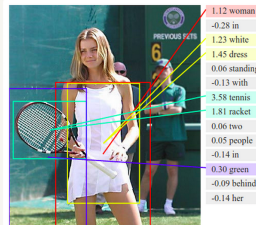
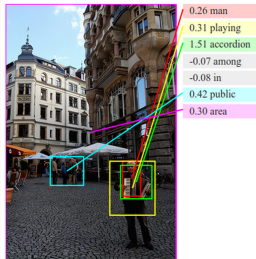
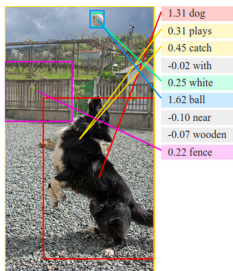


- Добавляя новые уровни, можно улучшить обобщающую способность, сделать признаки более общими:



СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- Более того, можно посмотреть, какие именно части картинки вызывают активации отдельных нейронов:

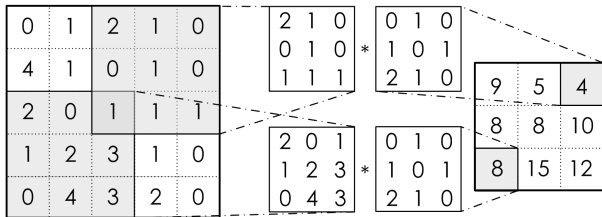


- Основная идея: мы применяем *одни и те же веса* по всей картинке.
- Это автоматически даёт устойчивость к переносам и т.п.
- И радикально сокращает число весов.
- Т.е. свёрточная структура – это такая радикальная форма регуляризации.

CNN ФОРМАЛЬНО

- Формально говоря, мы к каждому окну применяем одну и ту же маленькую матрицу.
- Если $\mathbf{x}^l$ – карта признаков на слое l , то двумерная свёртка размера $2d + 1$ с матрицей весов W размера $(2d + 1) \times (2d + 1)$ выглядит как

$$y_{i,j}^l = \sum_{-d \leq a, b \leq d} W_{a,b} x_{i+a, j+b}^l.$$



- Если $\mathbf{x}^l$ – карта признаков на слое l , то двумерная свёртка размера $2d + 1$ с матрицей весов W размера $(2d + 1) \times (2d + 1)$ выглядит как

$$y_{i,j}^l = \sum_{-d \leq a, b \leq d} W_{a,b} x_{i+a, j+b}^l.$$

- Затем обычно идёт нелинейность:

$$z_{i,j}^l = h(y_{i,j}^l).$$

- В современных сетях – почти всегда ReLU.

- И затем max-pooling (иногда average-pooling, но обычно max):

$$x_{i,j}^{l+1} = \max_{-d \leq a, b \leq d} z_{i+a, j+b}^l.$$

0	1	2	1
4	1	0	1
2	0	1	1
1	2	3	1

(a)

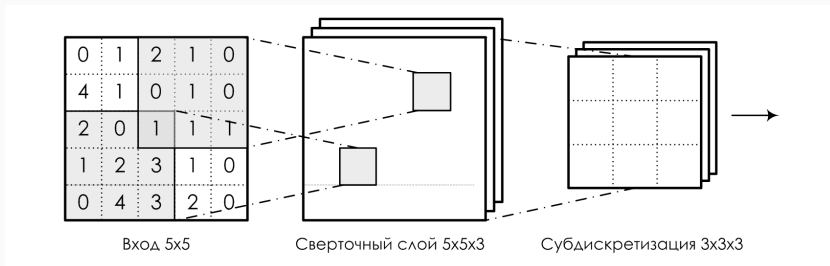
4	2	2
4	1	1
2	3	3

(б)

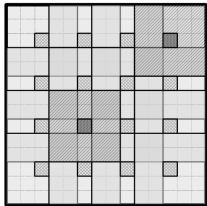
4	2
2	3

(в)

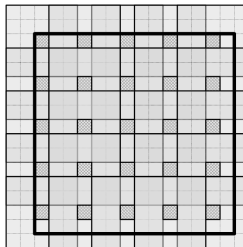
- Вот и получилась общая схема одного свёрточного слоя:



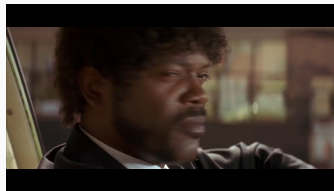
- Окнами можно покрывать по-разному; обычно всё-таки сохраняют исходный размер:



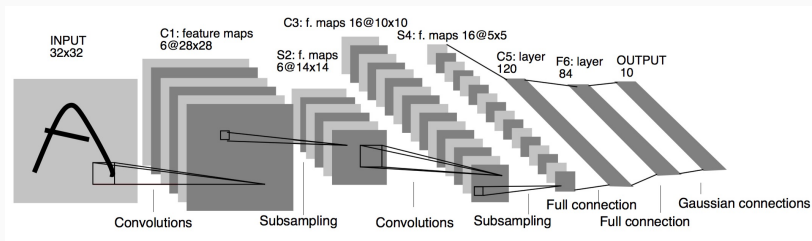
(а) окно 5x5 с шагом в 3 пикселя,
padding='VALID'



(б) окно 5x5 с шагом в 3 пикселя,
padding='SAME'



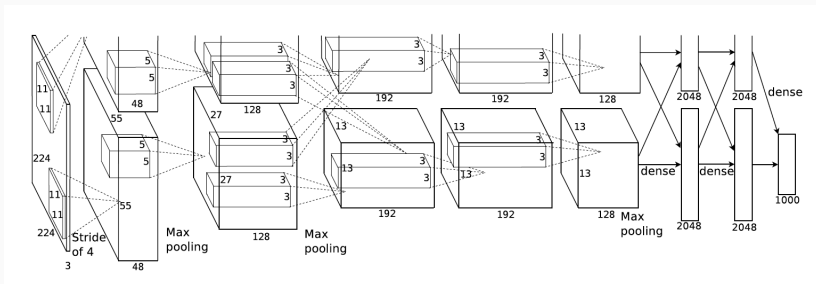
- Классическая архитектура *LeNet*:



- Современные CNN стали *очень* глубокими за счёт новых трюков – к ним мы скоро перейдём.

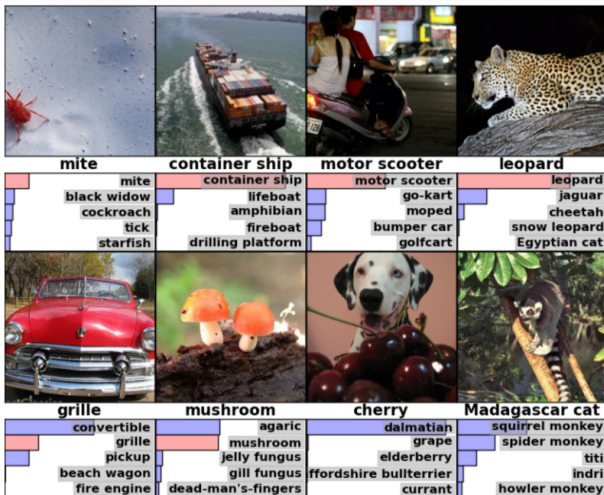
IMAGENET И ALEXNET

- ImageNet: > 15 миллионов картинок, ≈ 22000 категорий.
- Размечены руками (Amazon Mechanical Turk).
- Классическая сеть AlexNet (Krizhevsky, Sutskever, Hinton, 2012).



- С картинками можно делать data augmentation: изменять данные без изменения правильного ответа.
- (Krizhevsky, Sutskever, Hinton, 2012):
 - вырежем из 256×256 окна 224×224 (во время применения вырежем пять таких окон и усредним результаты);
 - отразим картинку по горизонтали;
 - поменяем интенсивности и освещённость (добавляя кратные главных компонент RGB-значений);
 - это увеличивает датасет в 2048 раз совершенно бесплатно. :)
- В AlexNet без data augmentation был бы оверфиттинг, это реально помогает.

ALEXNET HA IMAGENET



СПАСИБО!

Спасибо за внимание!

