

# СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

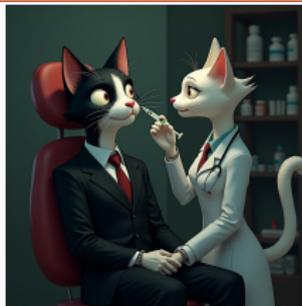
Сергей Николенко

СПбГУ — Санкт-Петербург

19 сентября 2024 г.

*Random facts:*

- 19 сентября — Международный пиратский день, точнее, International Talk Like a Pirate Day, учреждённый в 1995 году и отмечающийся пастафарианцами; aaarrrrgh, maties!
- 19 сентября 1648 г. в Квебеке Жак Буадон открыл первую в Канаде таверну, а 19 сентября 1654 г. был зарегистрирован первый в Канаде брак (невесте, кстати, было 11 лет)
- 19 сентября 1783 г. в присутствии короля Людовика XVI монгольфьер унёс из Версаля барана, петуха и утку на 4 километра
- 19 сентября 1888 в Спа (Бельгия) прошёл первый в истории конкурс красоты: 21 претендентка отбиралась по фотографиям, участницы конкурса жили в отдельном доме и приезжали на соревнования в закрытом экипаже; победила 18-летняя креолка из Гваделупы Берта Сукаре
- 19 сентября 1990 г. Ассоциация пользователей UNIX зарегистрировала домен верхнего уровня .su для применения на территории СССР
- 19 сентября 1991 г. два немецких туриста в седловине Хауслабьох в Тироле нашли Этци, ледяную мумию человека, жившего около 5300 лет назад; Этци было около 45 лет (много!), при нём были медный топор, каменный нож и 182-сантиметровый тисовый лук



## СВЁРТОЧНЫЕ СЕТИ: ИДЕЯ

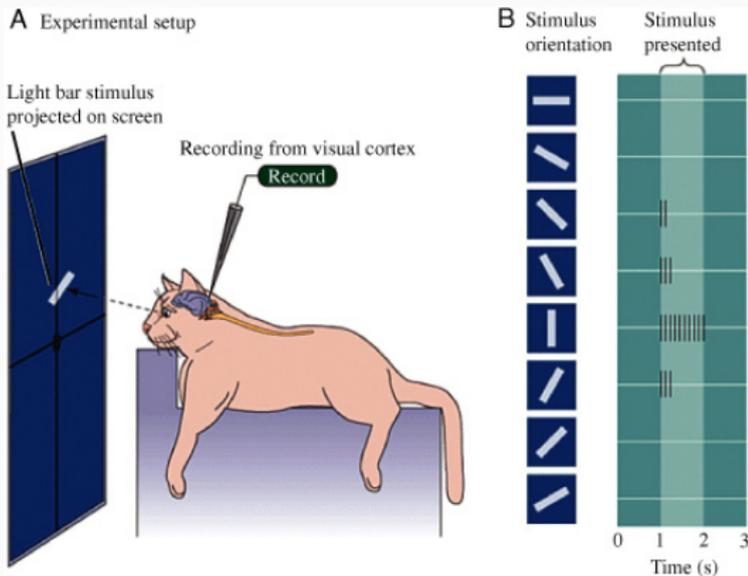
---

- 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.



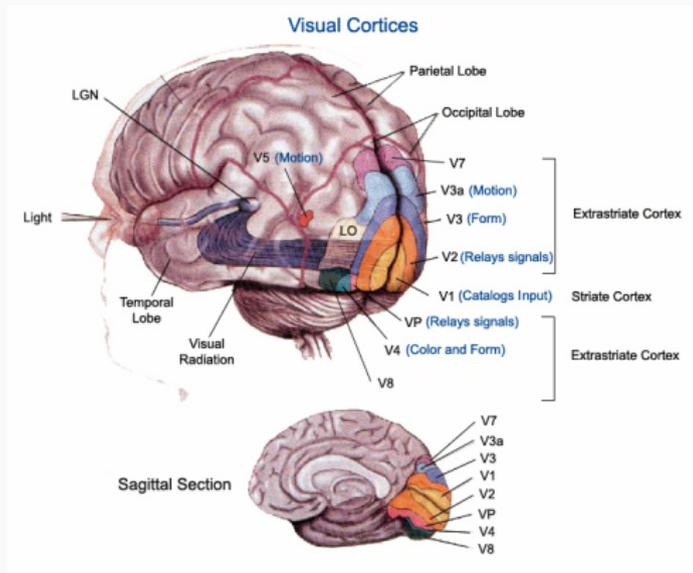
# ЗРИТЕЛЬНАЯ КОРА

- 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.

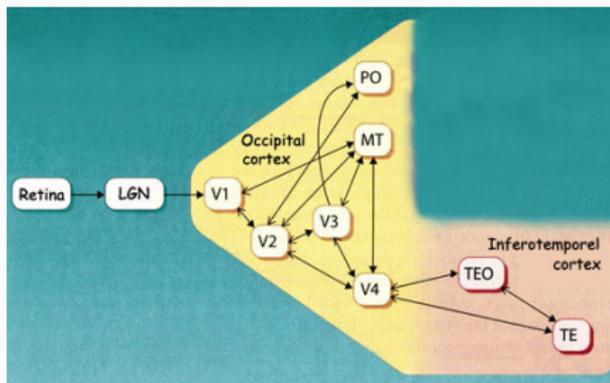


# ЗРИТЕЛЬНАЯ КОРА

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

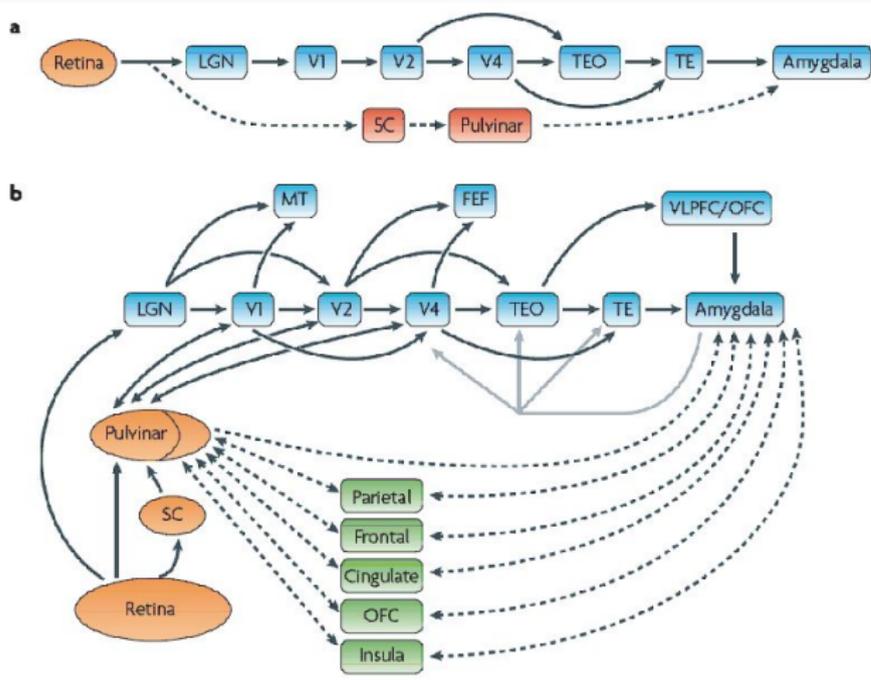


- Ну, вроде того. Мозг – сложная штука.



# ЗРИТЕЛЬНАЯ КОРА

- Правда сложная.

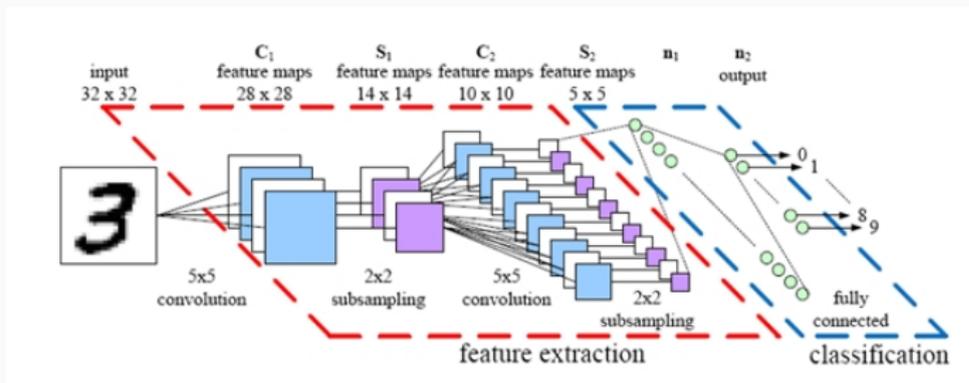




- Но некие условные уровни там действительно есть:
  - V1 – локальные признаки, топографические карты;
  - V2 – больше локальных признаков, бинокулярное зрение;
  - V3 – цвет, текстура, первые результаты сегментации;
  - V4 – геометрические формы, силуэты; это самый важный уровень для *внимания*;
  - 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.



- Разобьём изображение на окна:



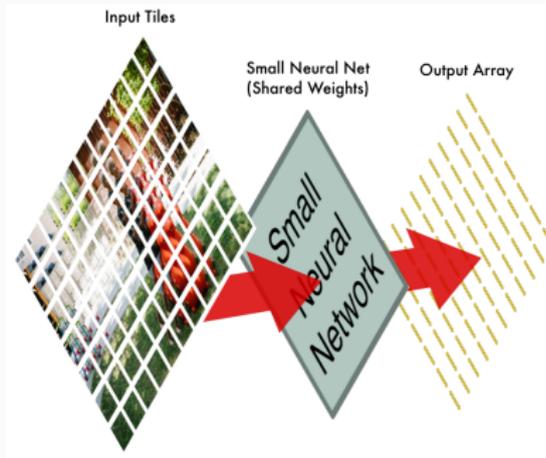
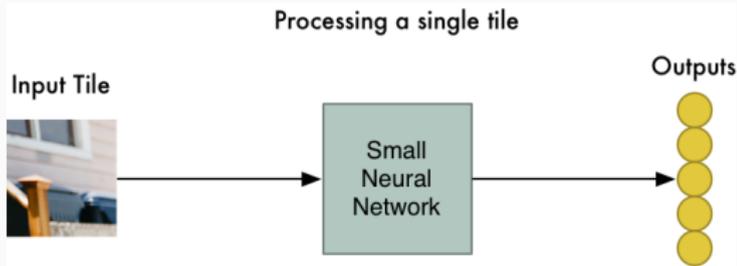
# СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- Разобьём изображение на окна:



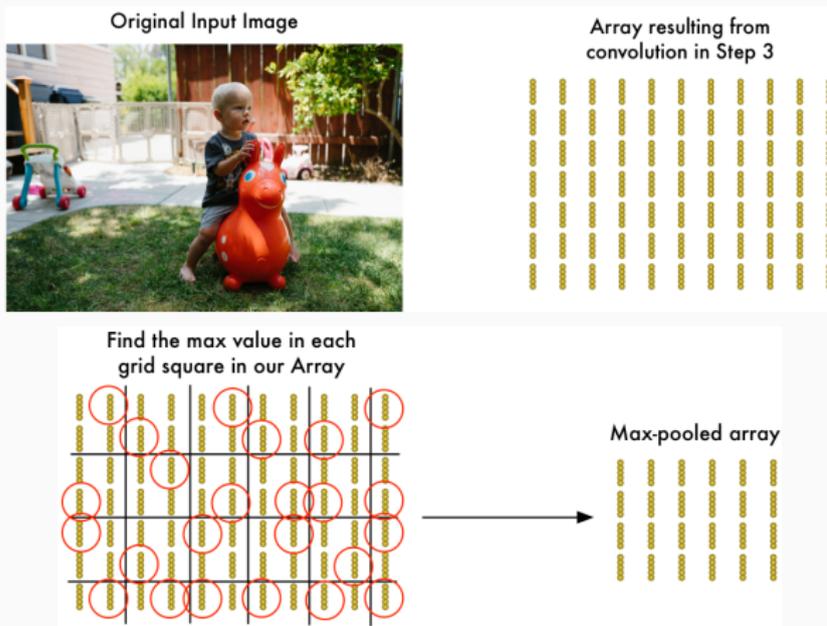
# СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

- Применим маленькую нейронную сеть к каждому окну:



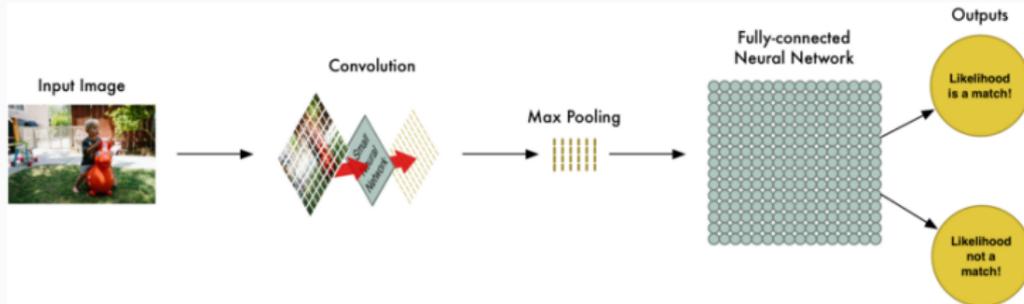
# СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

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

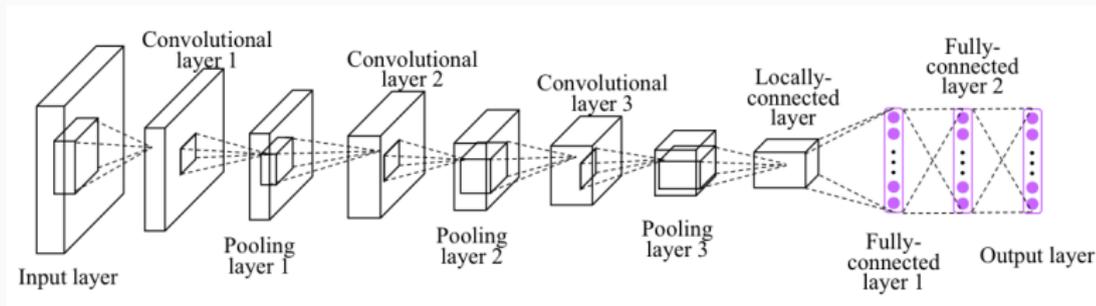


# СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

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

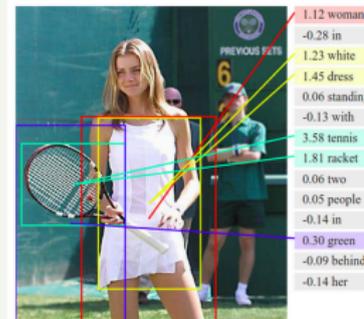
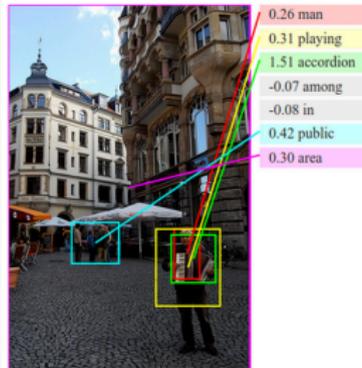
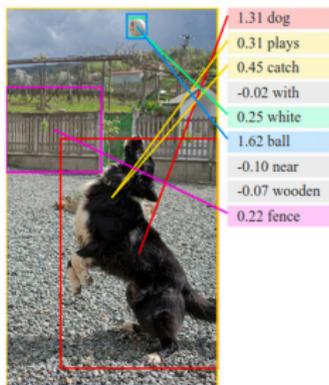


- Добавляя новые уровни, можно улучшить обобщающую способность, сделать признаки более общими:



# СВЁРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ: ИДЕЯ

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

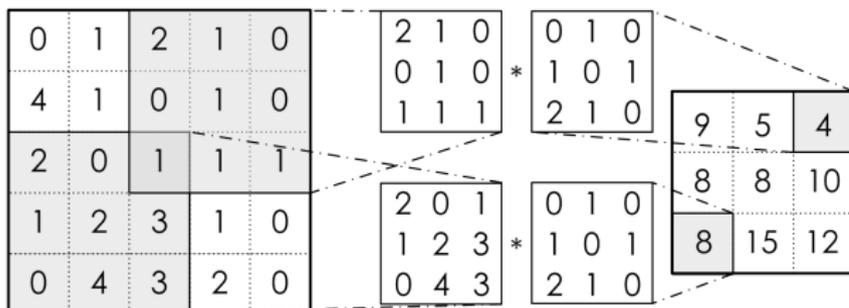


- Основная идея: мы применяем *одни и те же веса* по всей картинке.
- Это автоматически даёт устойчивость к переносам и т.п.
- И радикально сокращает число весов.
- Т.е. свёрточная структура – это такая радикальная форма регуляризации.

# 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)

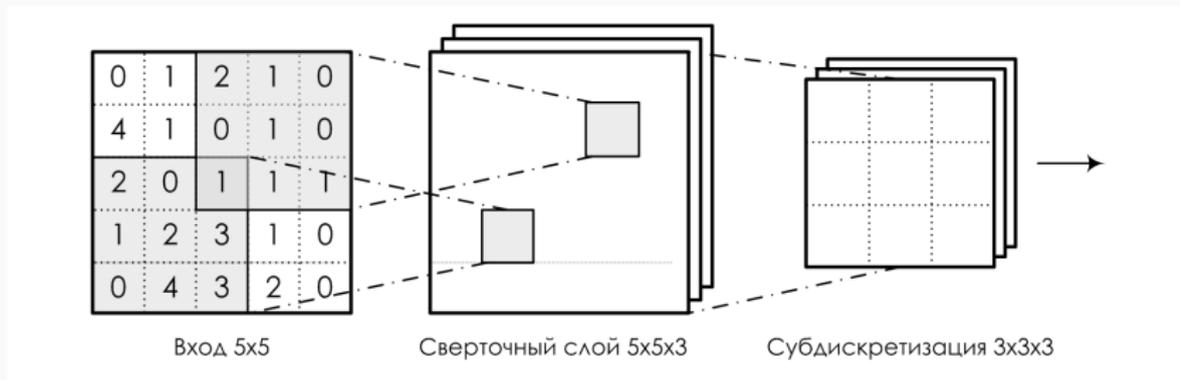
4	2	2
4	1	1
2	3	3

(б)

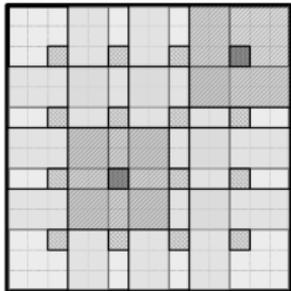
4	2
2	3

(в)

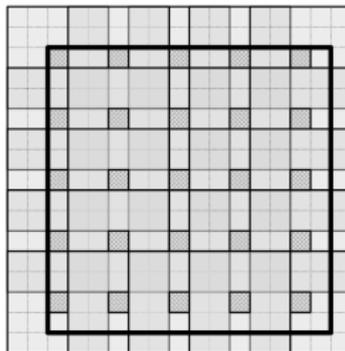
- Вот и получилась общая схема одного свёрточного слоя:



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



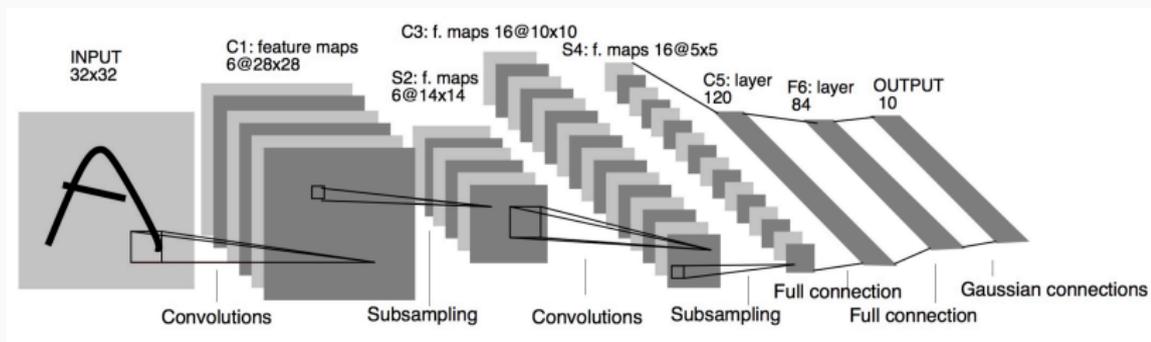
(а) окно 5x5 с шагом в 3 пикселя,  
padding='VALID'



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



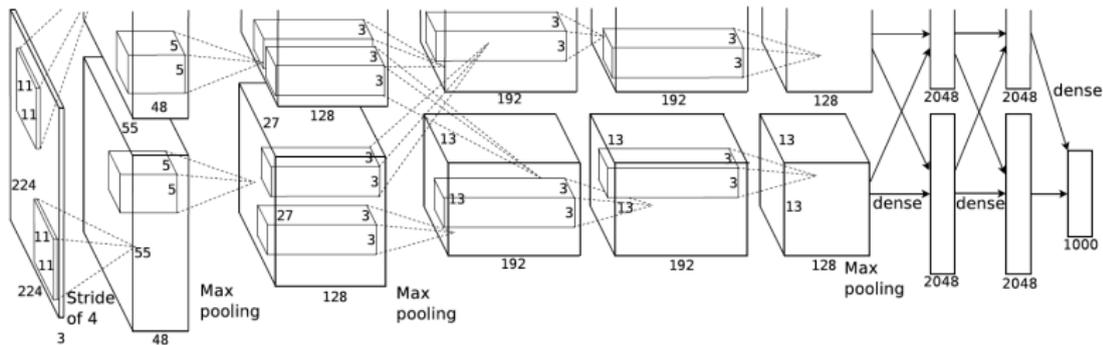
- Классическая архитектура *LeNet*:



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

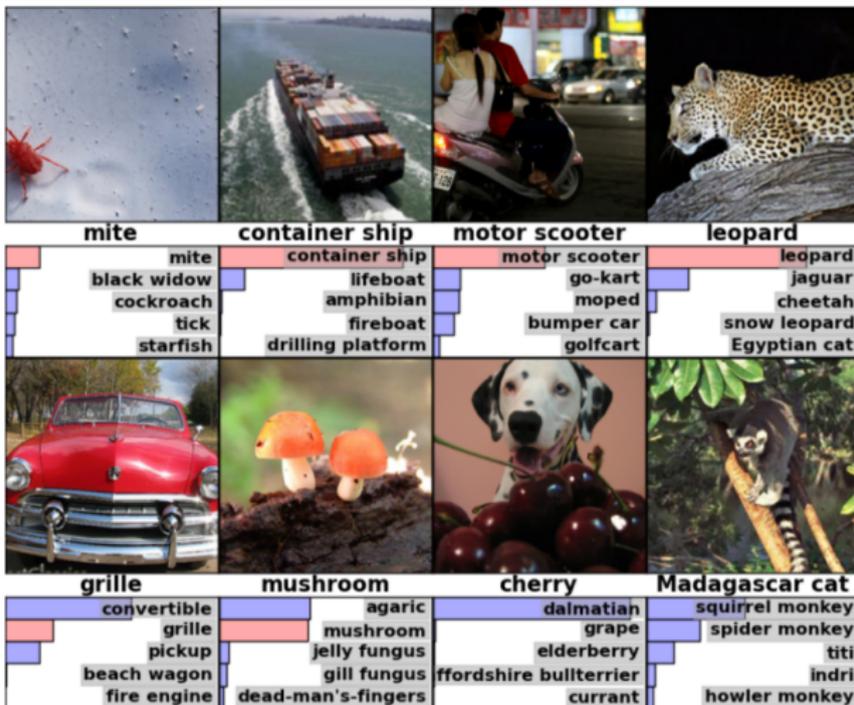
# IMAGENET И ALEXNET

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

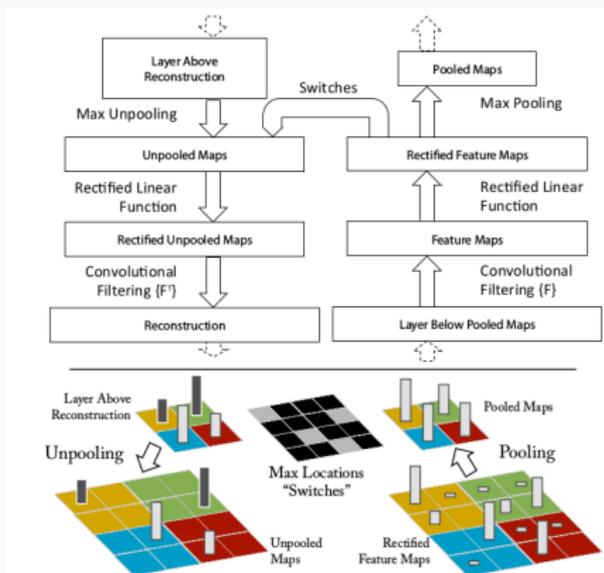


- С картинками можно делать data augmentation: изменять данные без изменения правильного ответа.
- (Krizhevsky, Sutskever, Hinton, 2012):
  - вырежем из  $256 \times 256$  окна  $224 \times 224$  (во время применения вырежем пять таких окон и усредним результаты);
  - отразим картинку по горизонтали;
  - 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.

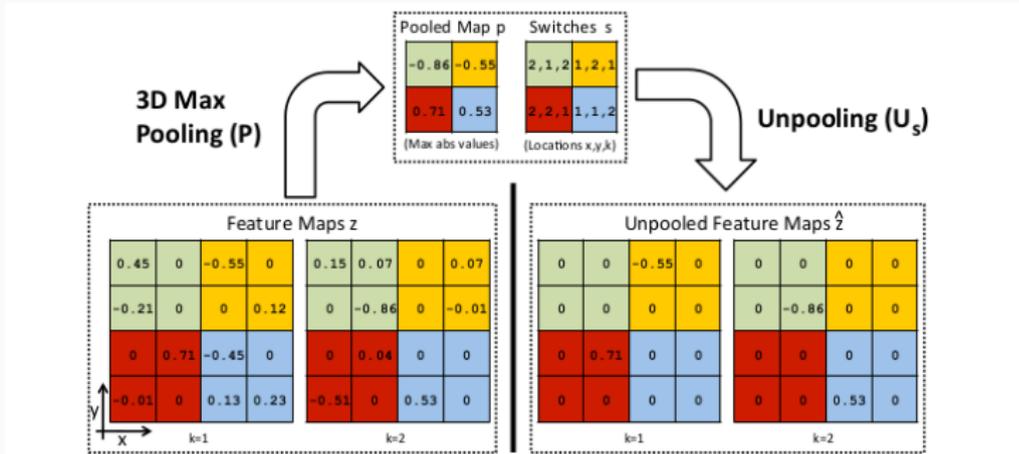
# ALEXNET HA IMAGENET



- (Zeiler et al., 2011; Zeiler, Fergus, 2014): деконволюция как обратная свёртке операция.
- Чтобы обратить свёртку, достаточно транспонировать.

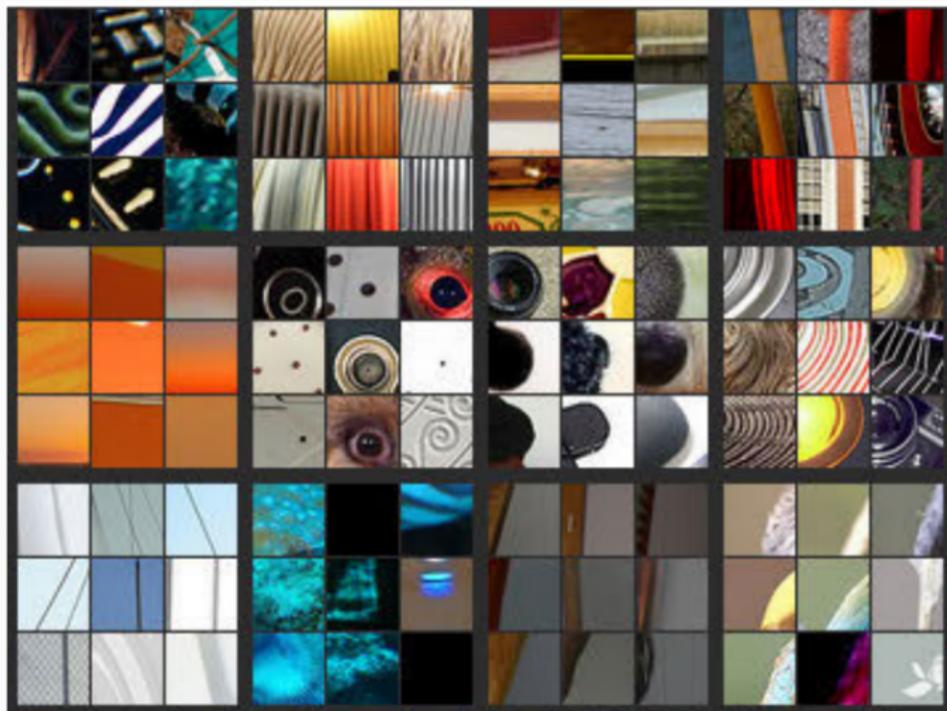


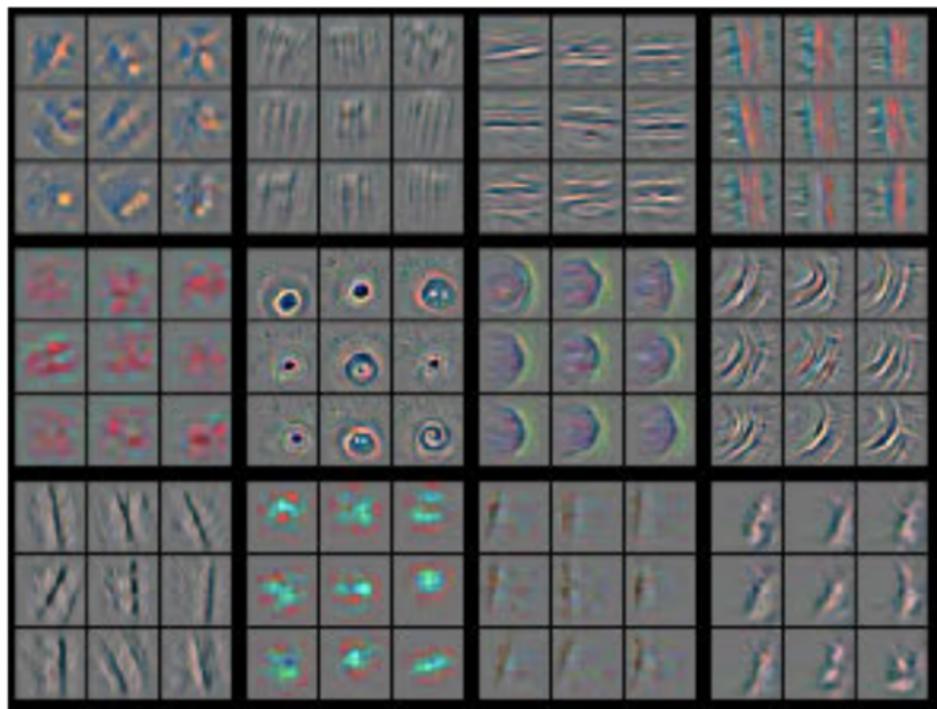
- Интересный вопрос – как обратить субдискретизацию.
- Она, конечно, необратима, но можно запомнить положение максимумов и обратить примерно.



(Zeiler, Fergus, 2014): визуализируем признаки через деконволюцию. Это сеть, обученная на ImageNet.

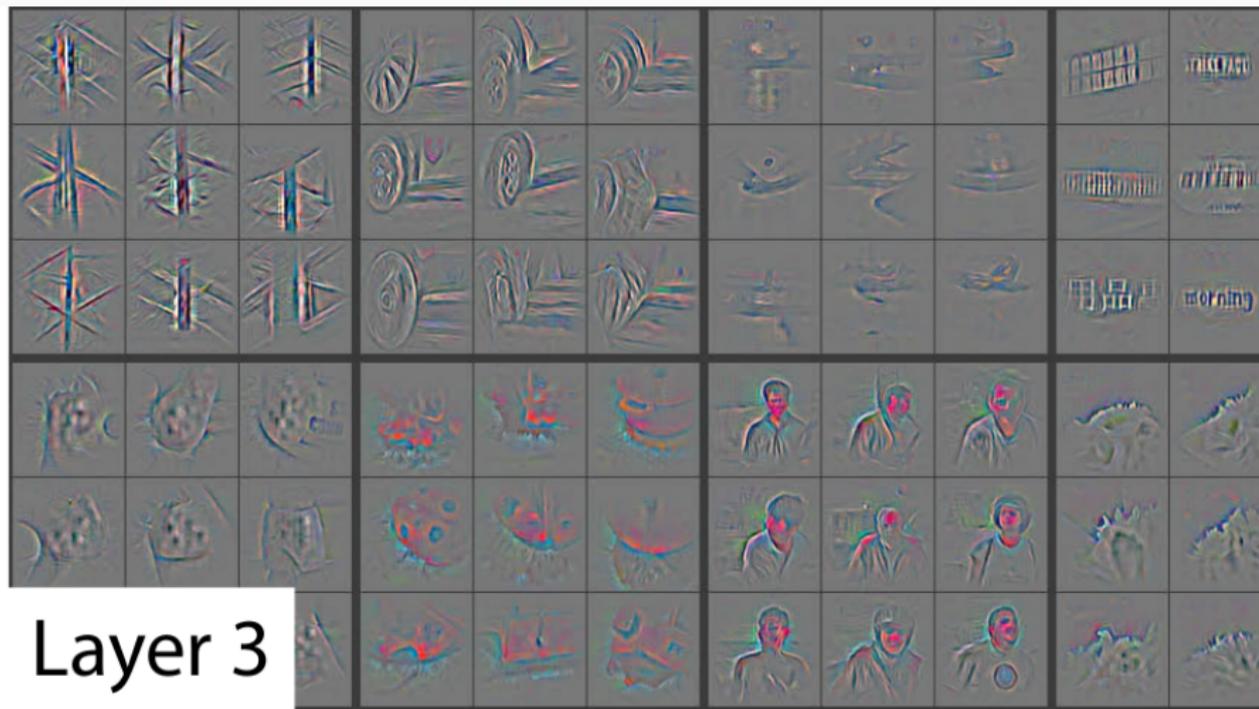


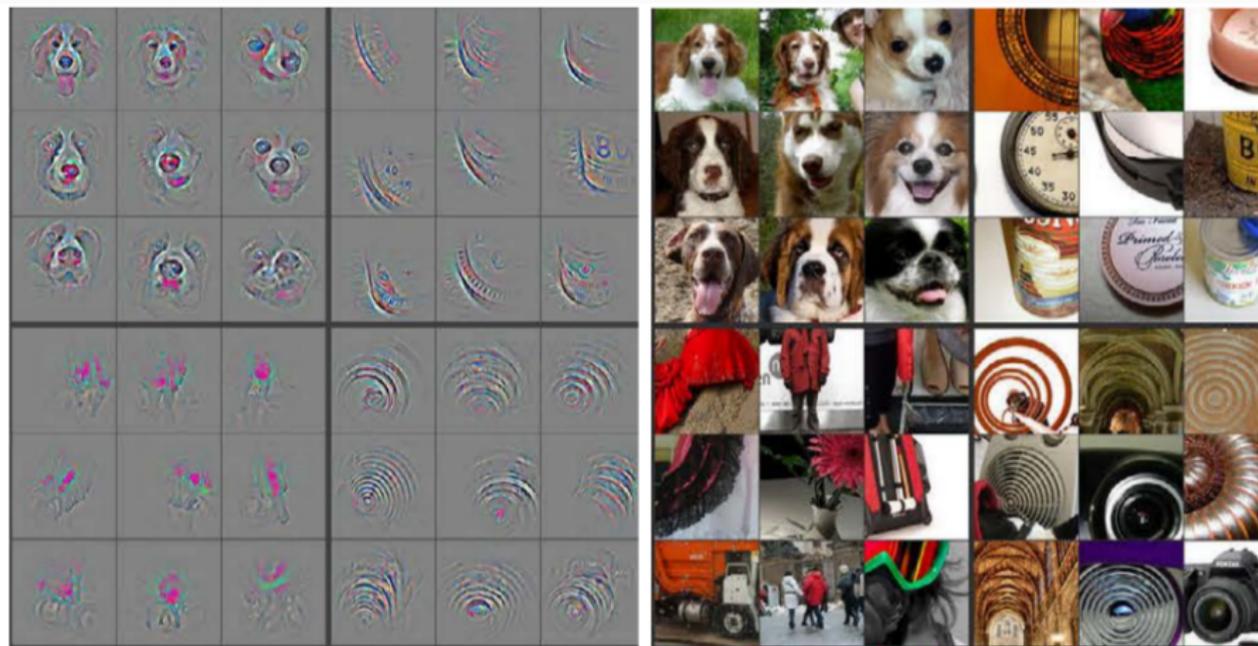




# (ZEILER, FERGUS, 2014): УРОВЕНЬ 3

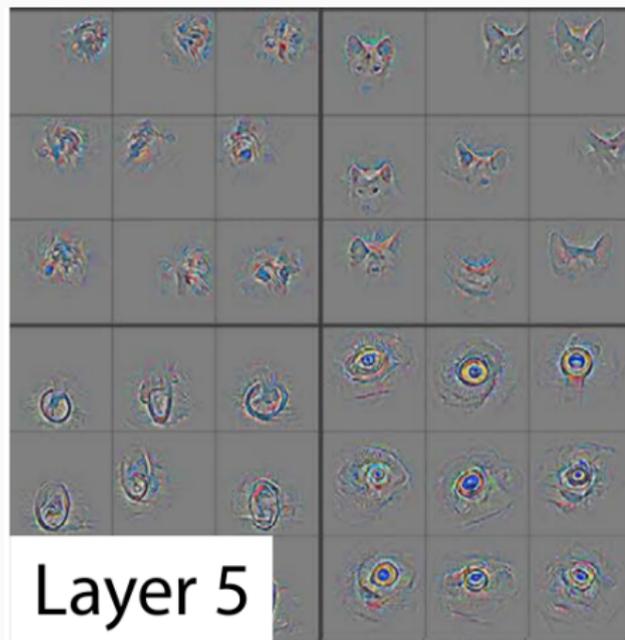












# СОВРЕМЕННЫЕ СВЁРТОЧНЫЕ АРХИТЕКТУРЫ

---

- Разных моделей, использующих CNN, очень много, и всё время появляются новые.
- Всё можно скачать, часто даже уже веса, предобученные на стандартных датасетах, есть.
- Но обычно они основаны на нескольких относительно стандартных идеях. Их-то мы и рассмотрим.

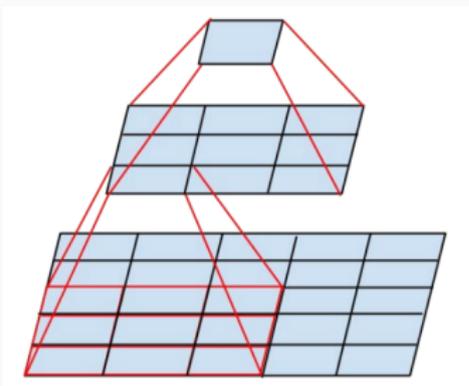
## Model Zoo

Discover open source deep learning code and pretrained models.

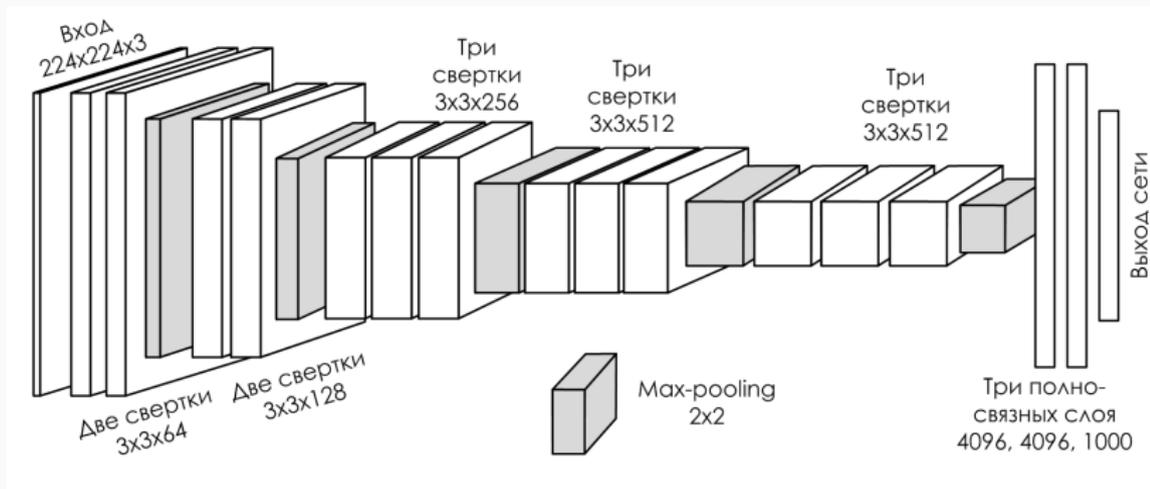
[Browse Frameworks](#)

[Browse Categories](#)

- VGG (Oxford Visual Geometry Group): давайте представлять большие свёртки как комбинации свёрток  $3 \times 3$ .
- Это уменьшает число весов и делает сеть глубже.

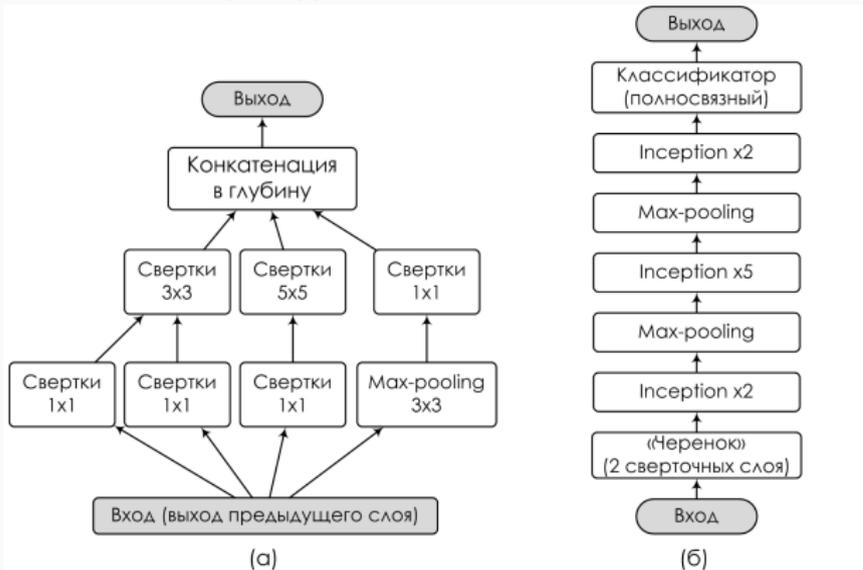


- Сеть VGG, использованная в ImageNet Large Scale Visual Recognition Competition (ILSVRC-2014).

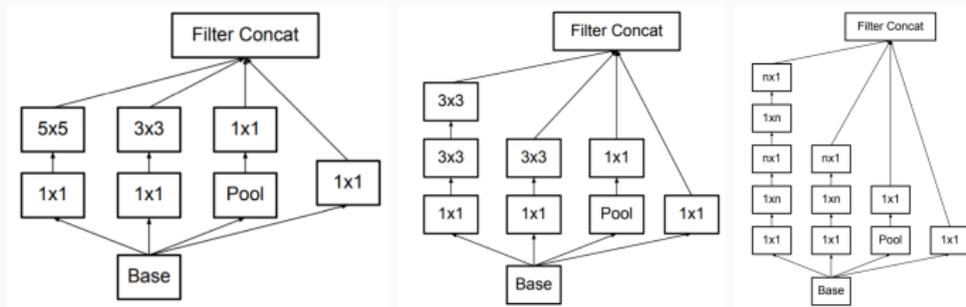
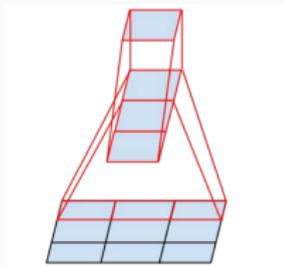


- А сейчас NVIDIA специально оптимизирует GPU для свёрток  $3 \times 3$ .

- *Inception*, разработанная командой *GoogLeNet*: давайте используем идею “сеть в сети” (network in network), чтобы сократить веса ещё больше!
- И дополнительные классификаторы (просто лишние слагаемые в целевую функцию).



- Во второй версии (Szegedy et al., 2015) свёртки  $n \times n$  заменили на комбинации свёрток  $n \times 1$  и  $1 \times n$ :





- Label smoothing:

- в обычном классификаторе мы обучаем результаты softmax с целевой переменной  $q(k) = [k = y]$ , т.е. аргументы softmax хотят расти неограниченно;
- модели становятся слишком уверены в себе, оверфиттинг;
- решение: давайте вместо этого оптимизировать

$$q'(k) = (1 - \epsilon)[k = y] + \epsilon u(k)$$

для какого-то априорного  $u(k)$ , например  $u(k) = \frac{1}{K}$ .

Network	Crops Evaluated	Top-5 Error	Top-1 Error
GoogLeNet [20]	10	-	9.15%
GoogLeNet [20]	144	-	7.89%
VGG [18]	-	24.4%	6.8%
BN-Inception [7]	144	22%	5.82%
PReLU [6]	10	24.27%	7.38%
PReLU [6]	-	21.59%	5.71%
Inception-v3	12	19.47%	4.48%
Inception-v3	144	<b>18.77%</b>	<b>4.2%</b>

СПАСИБО!

Спасибо за внимание!

