СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

Сергей Николенко

СПбГУ — Санкт-Петербург

04 сентября 2025 г.

Random facts:

- 4 сентября 476 г. предводитель варваров-наёмников Одоакр сверг римского императора Августула; так закончилась Западная Римская империя и началось королевство Одоакра, которое не просуществовало и двадцати лет
- 4 сентября 1607 г. произошло «бегство графов»: 99 ирландских аристократов, которые были потомками ирландских королей и вождей кланов, сбежали из Ирландии на кораблях в Испанию; из-за шторма они высадились во Франции, а затем направились в Рим, где получили убежище, но в Ирландию уже не вернулись
- 4 сентября 1837 г. в Нью-Йоркском университете Сэмюэл Морзе впервые продемонстрировал телеграф
- 4 сентября 1884 г. в ходе контрреформ Александра III в России была отменена автономия университетов
- 4 сентября 1975 г. в телеэфир вышел первый выпуск игры «Что? Где? Когда?»; ведущего не было, две семьи играли друг против друга, и раунды снимались у них дома
- 4 сентября 1998 г. была основана компания Google; она была расположена в Менло-Парке, в гараже Сьюзен Воджиски, подруги Ларри Пейджа и Сергея Брина

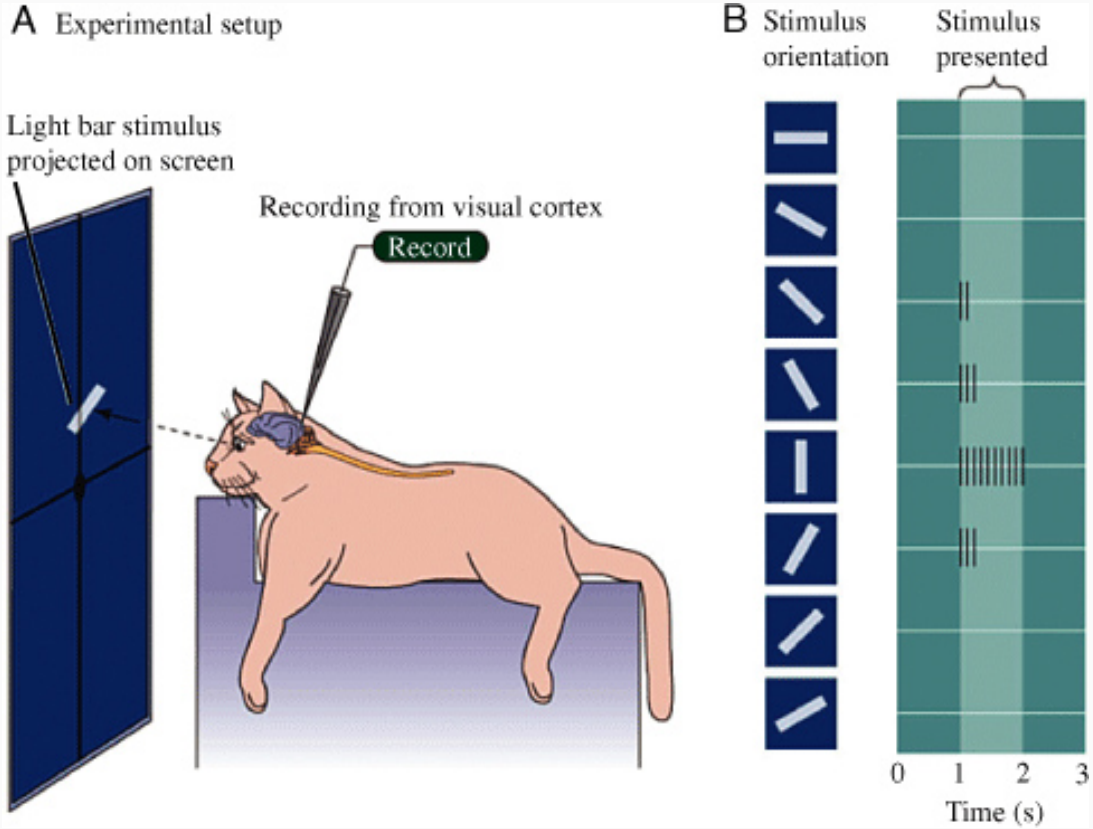
Свёрточные сети: идея

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.

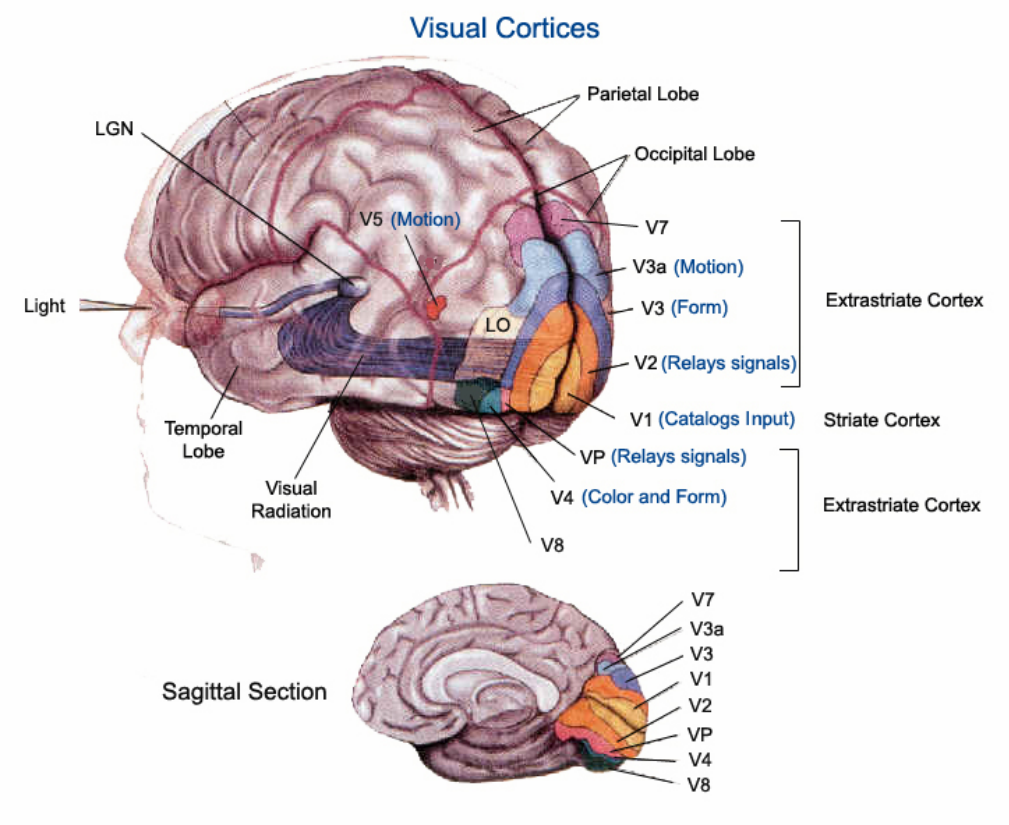
Зрительная кора

- 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.



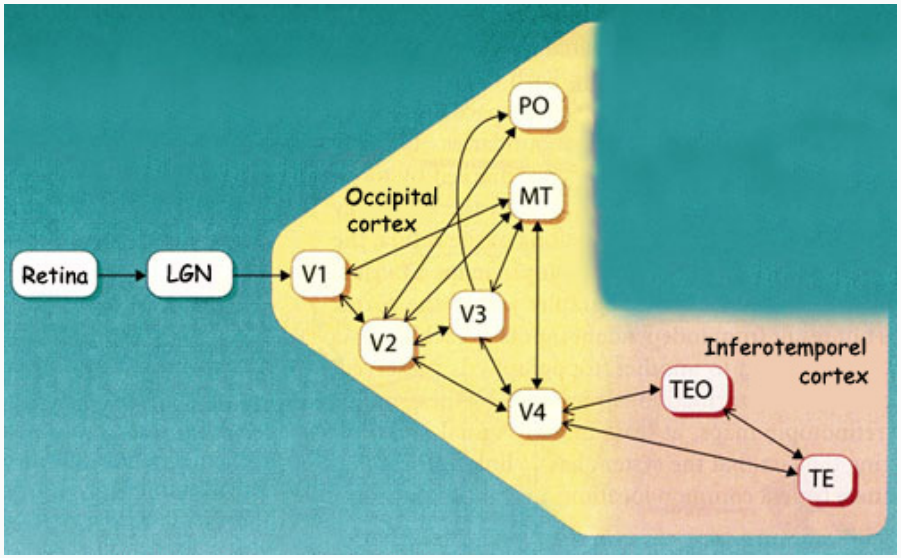
Зрительная кора

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



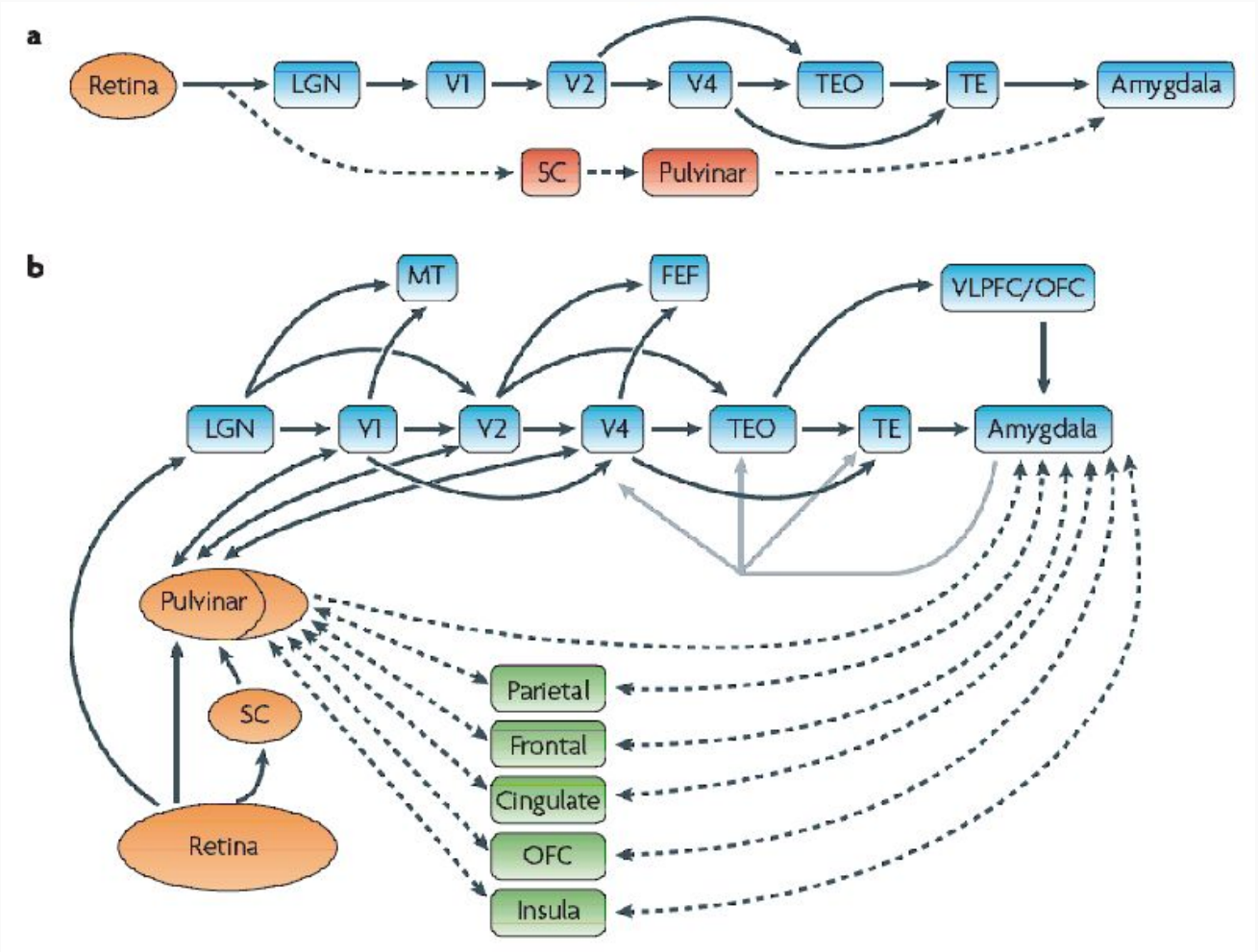
Зрительная кора

- Ну, вроде того. Мозг – сложная штука.



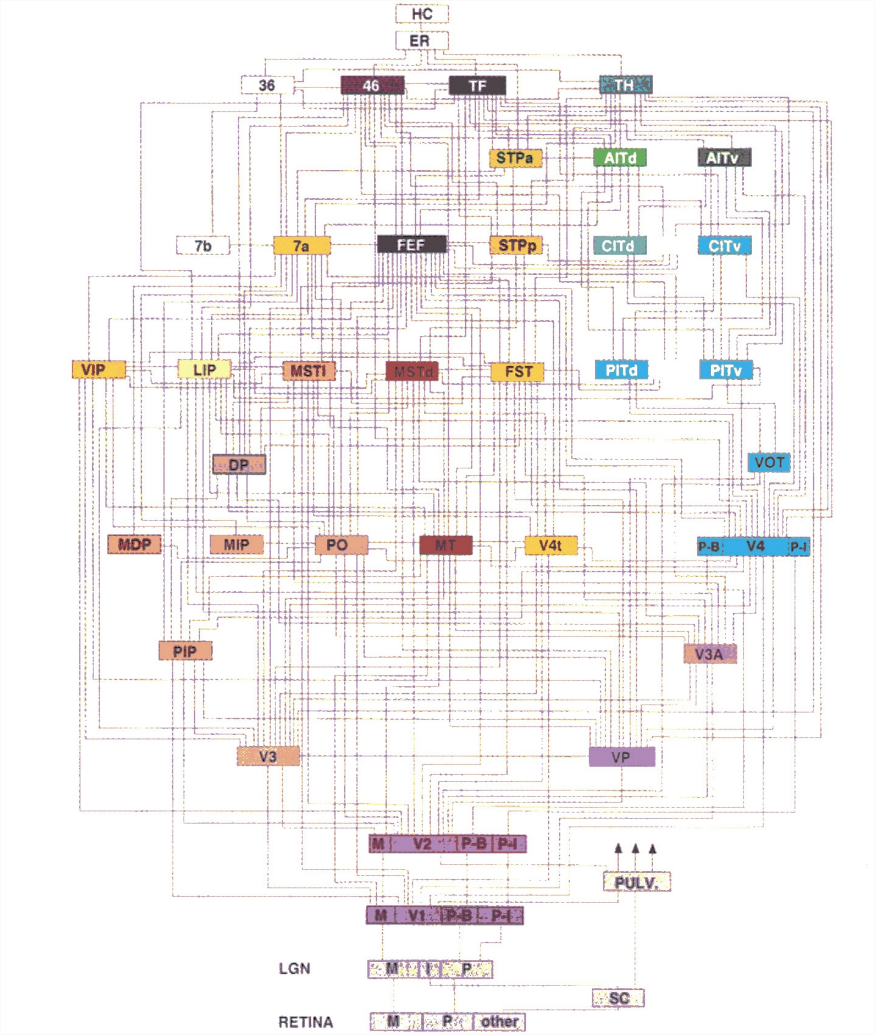
Зрительная кора

- Правда сложная.



Зрительная кора

- Честно-честно.

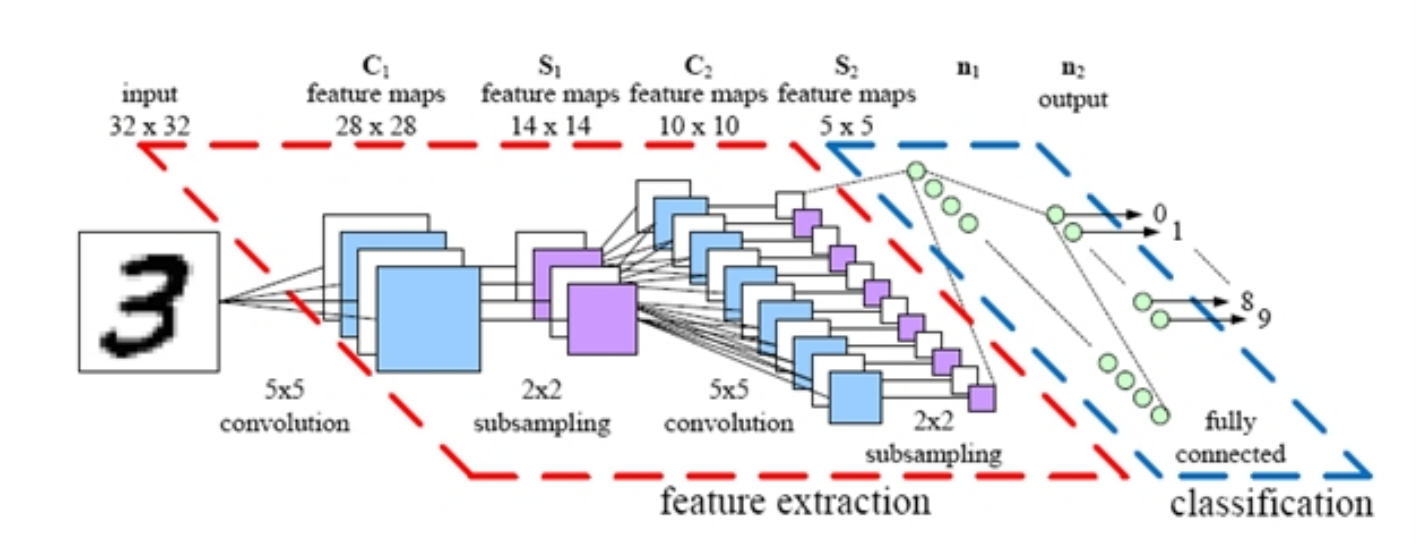


Зрительная кора

- 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.

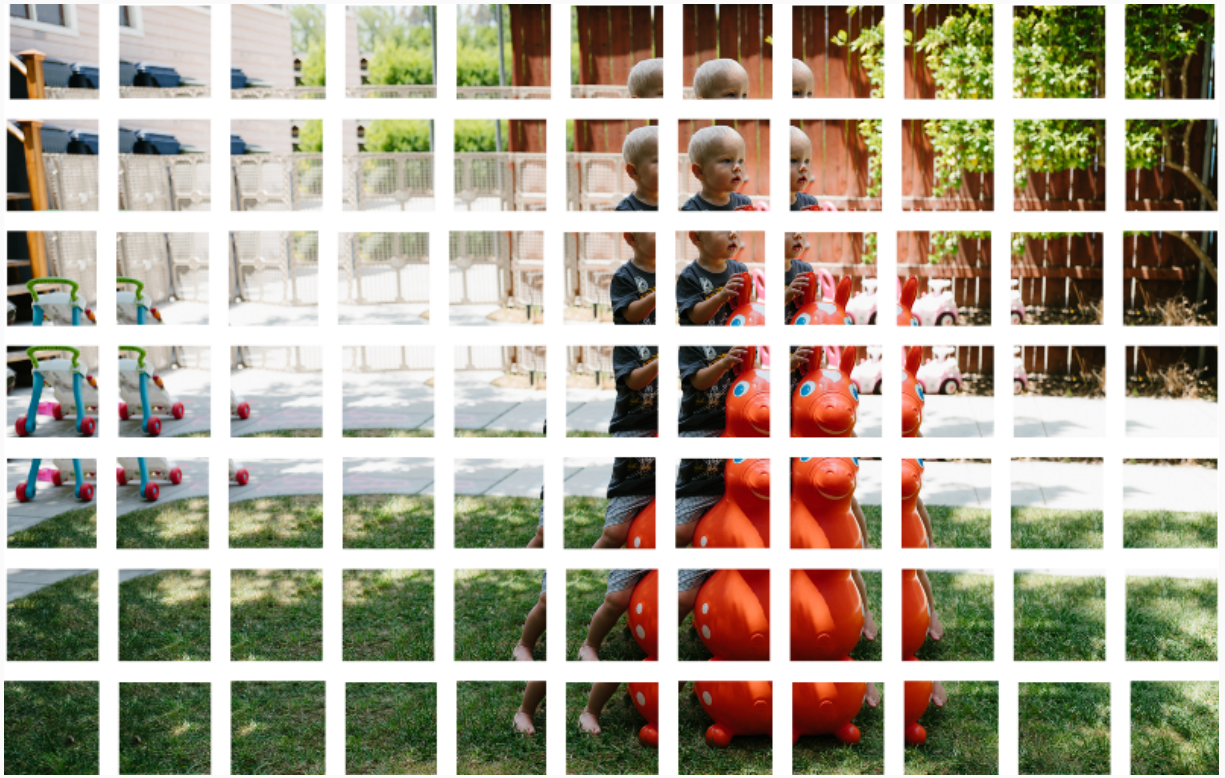


Свёрточные нейронные сети: идея

- Разобьём изображение на окна:

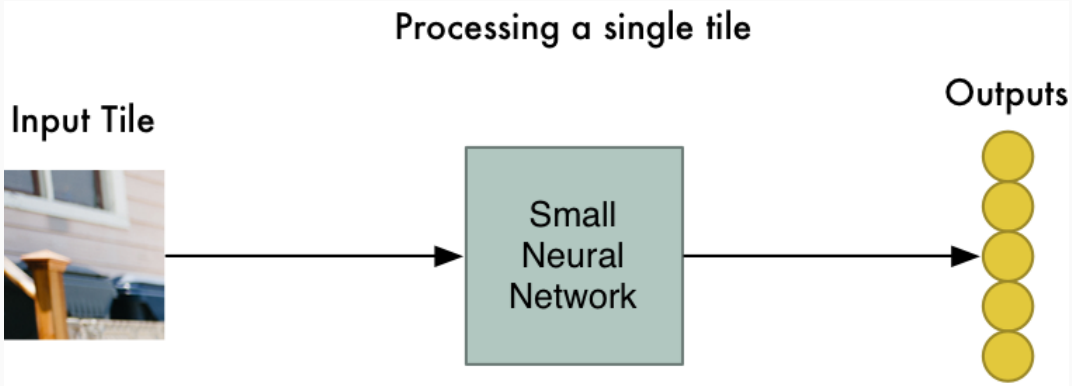
Свёрточные нейронные сети: идея

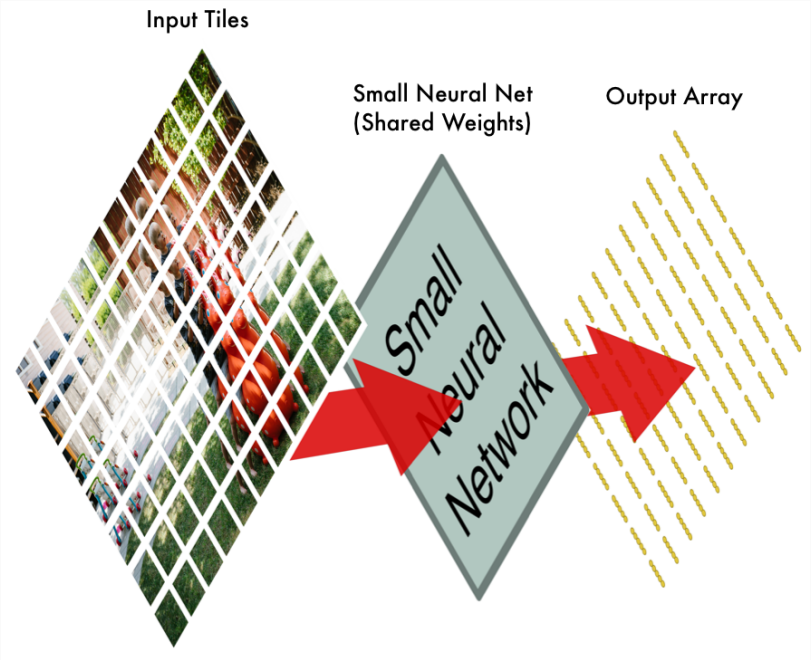
- Разобьём изображение на окна:



Свёрточные нейронные сети: идея

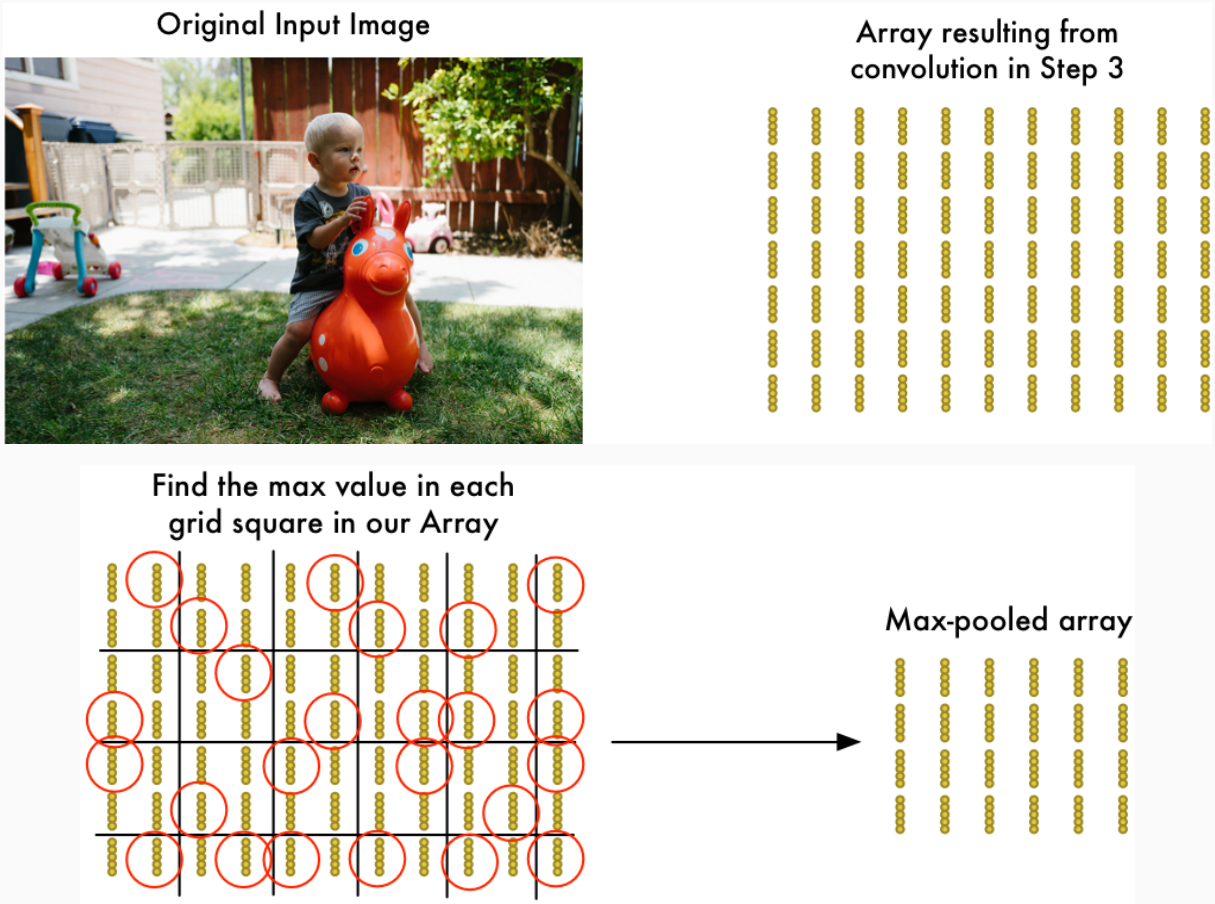
- Применим маленькую нейронную сеть к каждому окну:





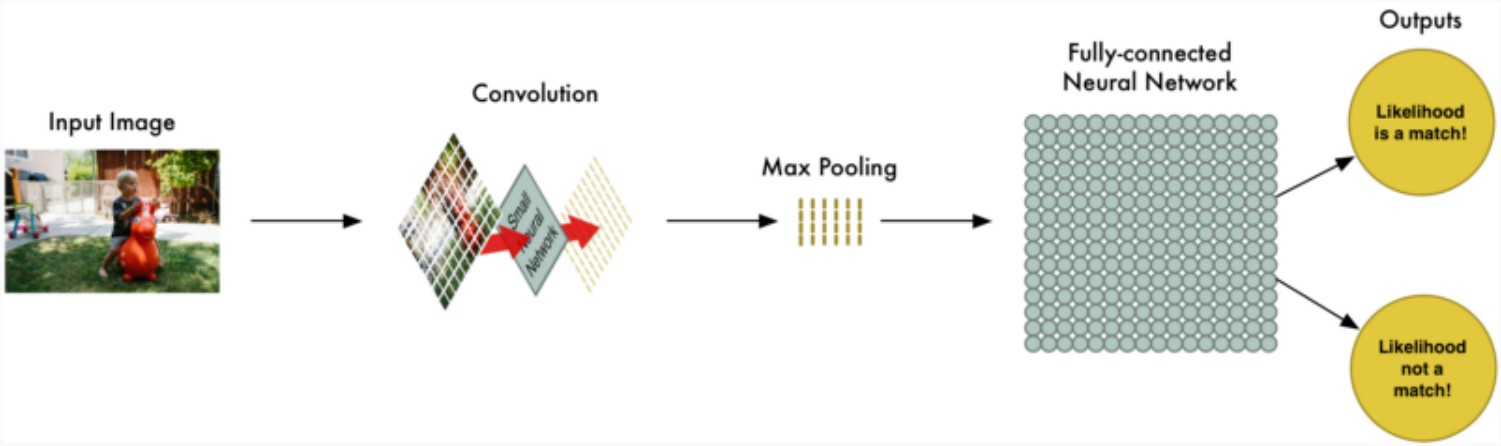
Свёрточные нейронные сети: идея

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

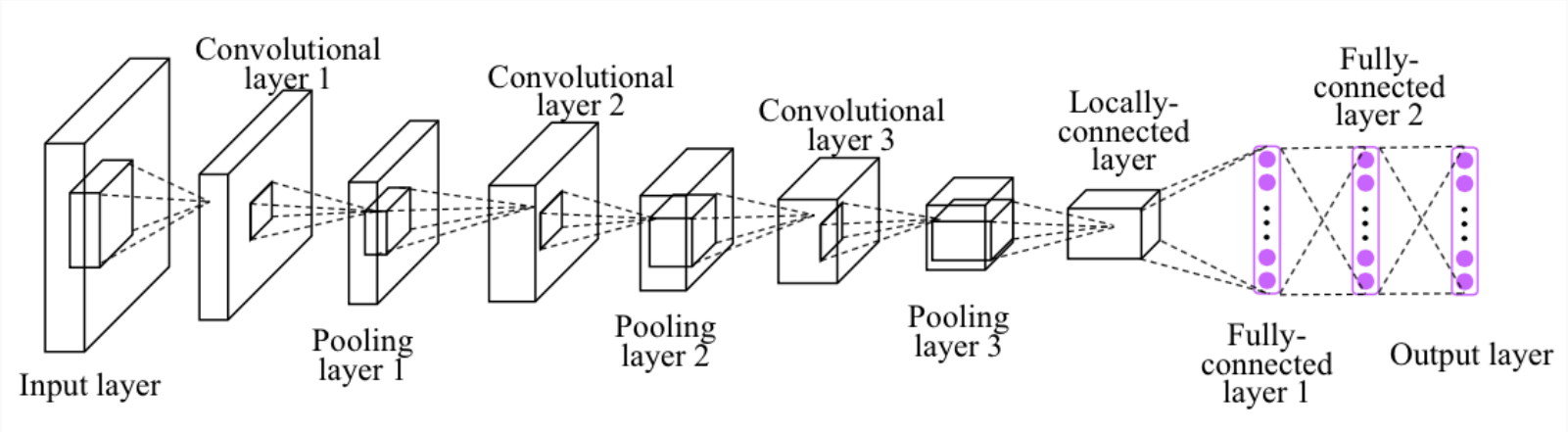


Свёрточные нейронные сети: идея

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

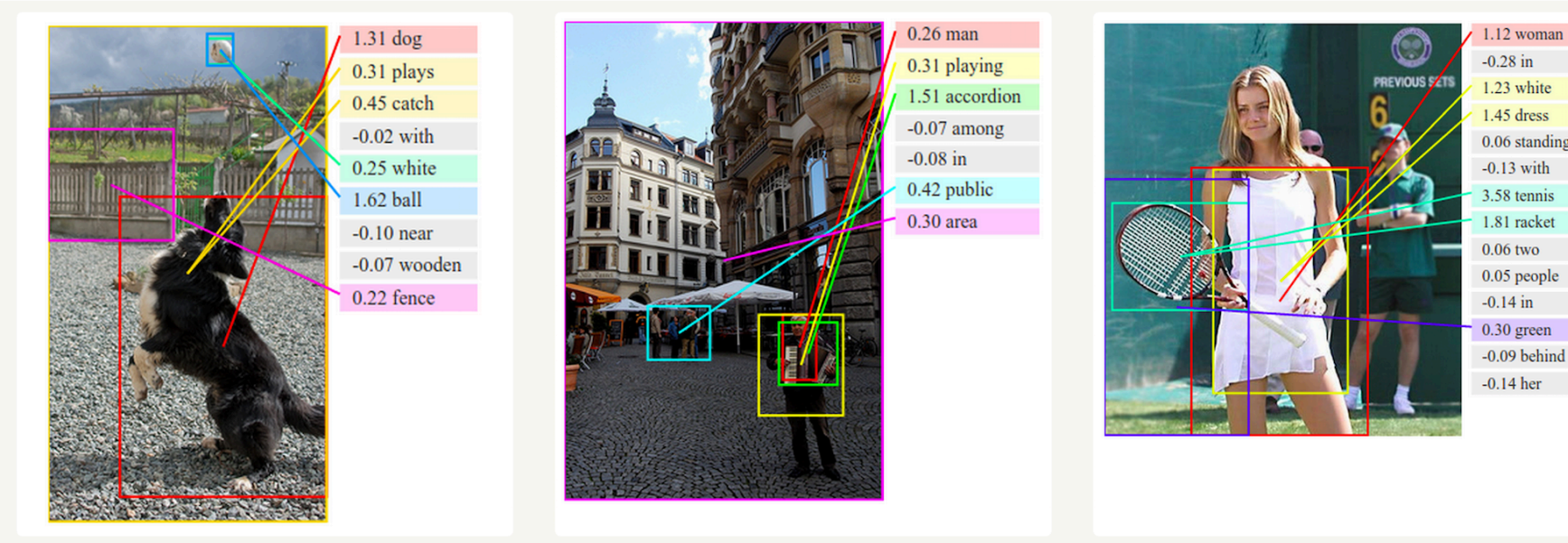


- Добавляя новые уровни, можно улучшить обобщающую способность, сделать признаки более общими:



Свёрточные нейронные сети: идея

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



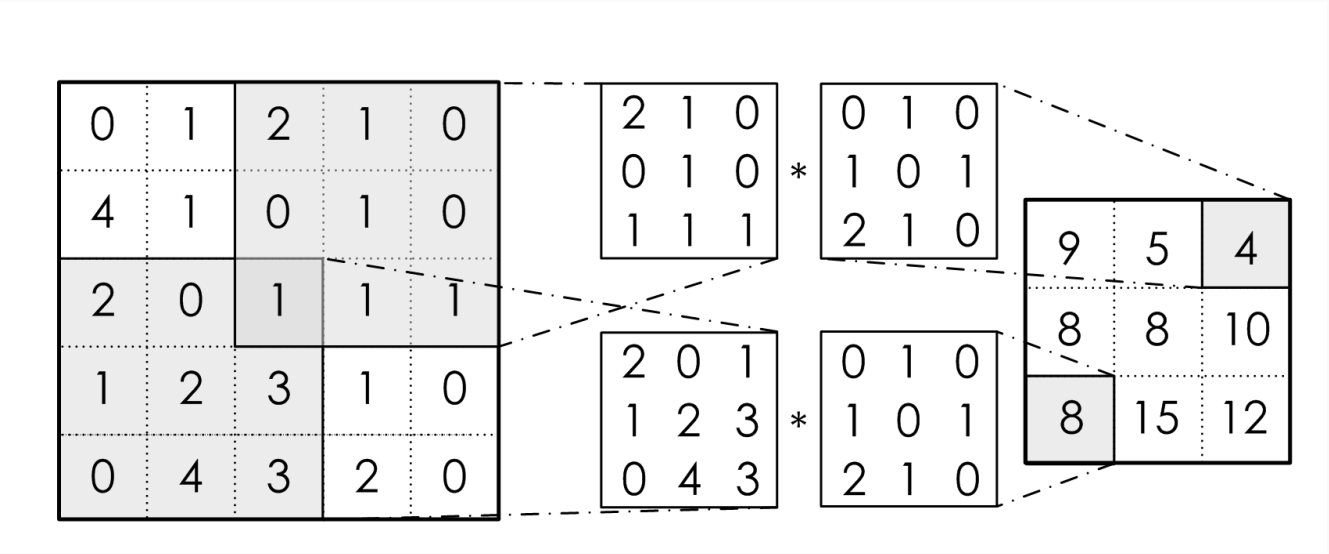
Свёрточные нейронные сети: идея

- 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

$$y^l_{i,j} = \sum_{-d<=a,b<=d} W_{a,b} x^l_{i+a,j+b}.$$



CNN формально

- 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

$$y^l_{i,j} = \sum_{-d<=a,b<=d} W_{a,b} x^l_{i+a,j+b}.$$

- Затем обычно идёт нелинейность:

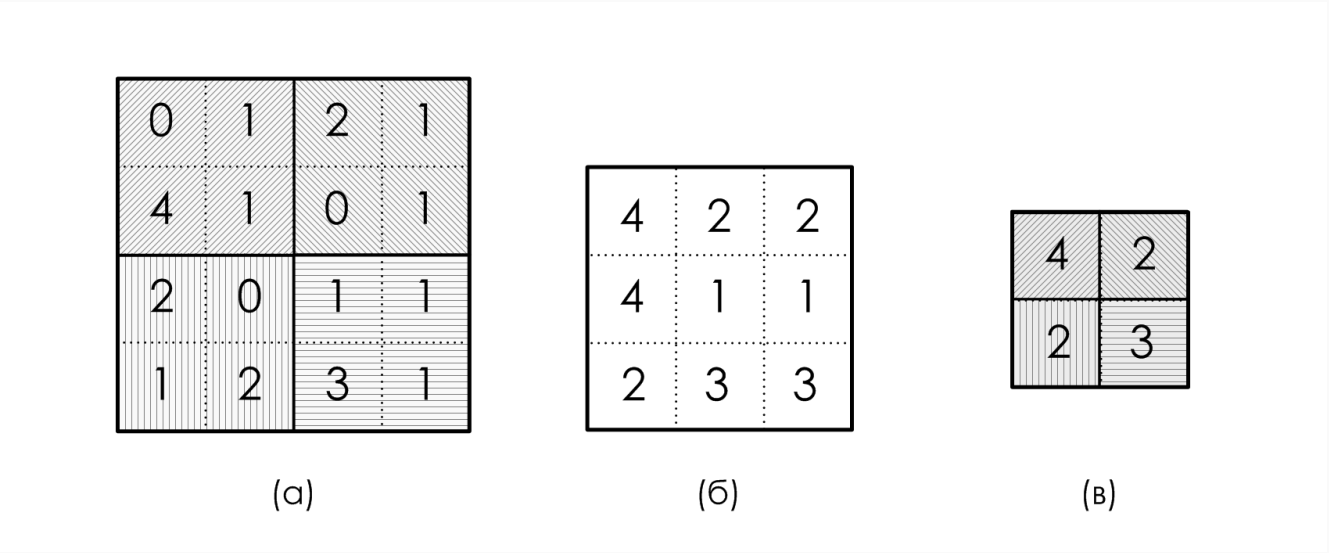
$$z^l_{i,j} = h(y^l_{i,j}).$$

- В современных сетях – почти всегда ReLU.

CNN формально

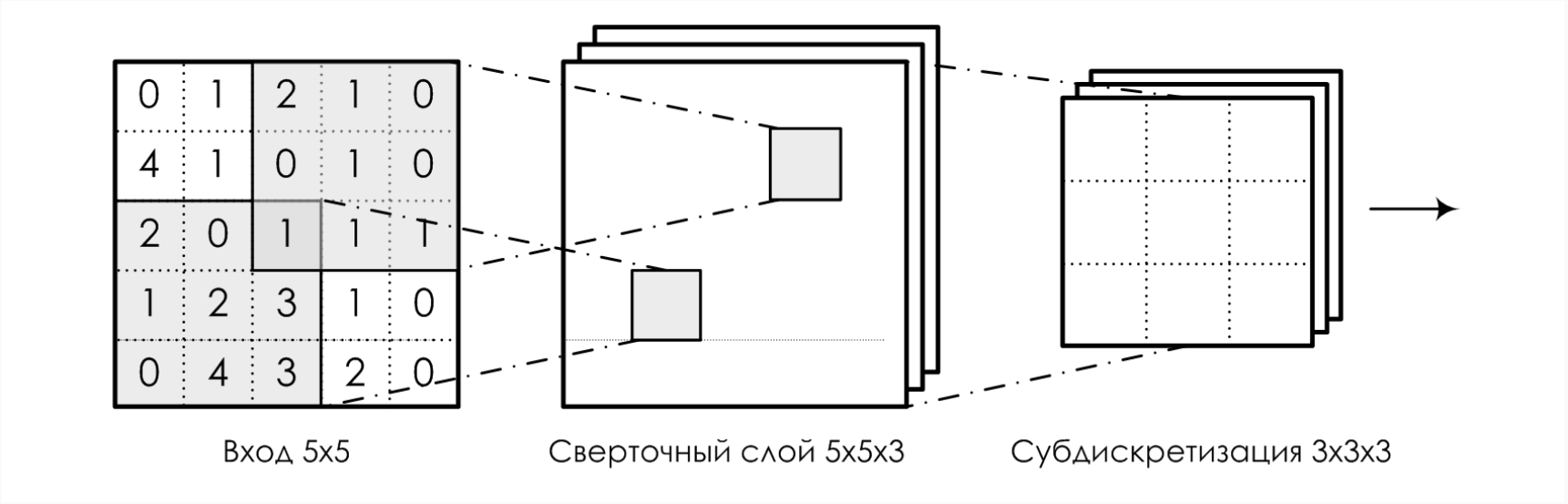
- И затем max-pooling (иногда average-pooling, но обычно max):

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



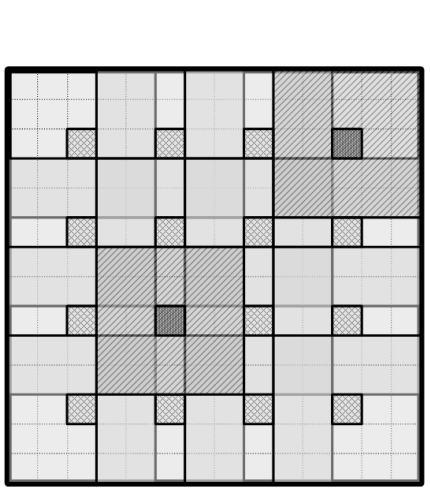
CNN формально

- Вот и получилась общая схема одного свёрточного слоя:

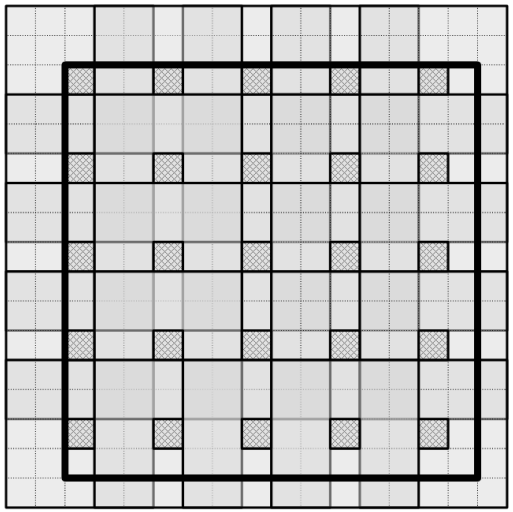


CNN формально

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



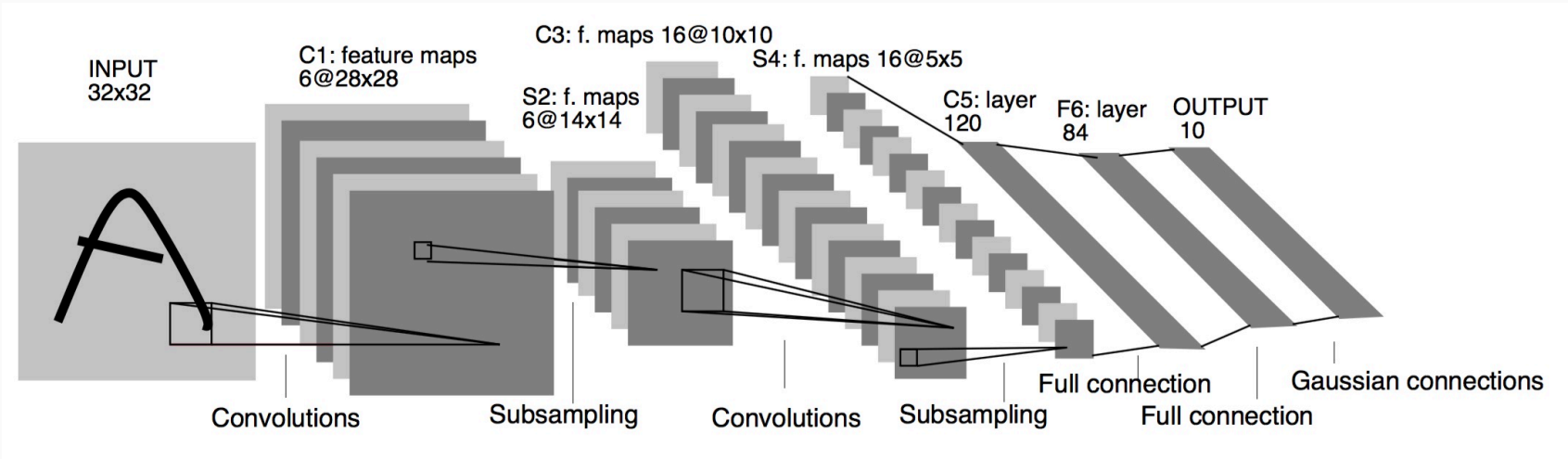
(а) окна 5x5 с шагом в 3 пикселя, padding='VALID'



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

CNN формально

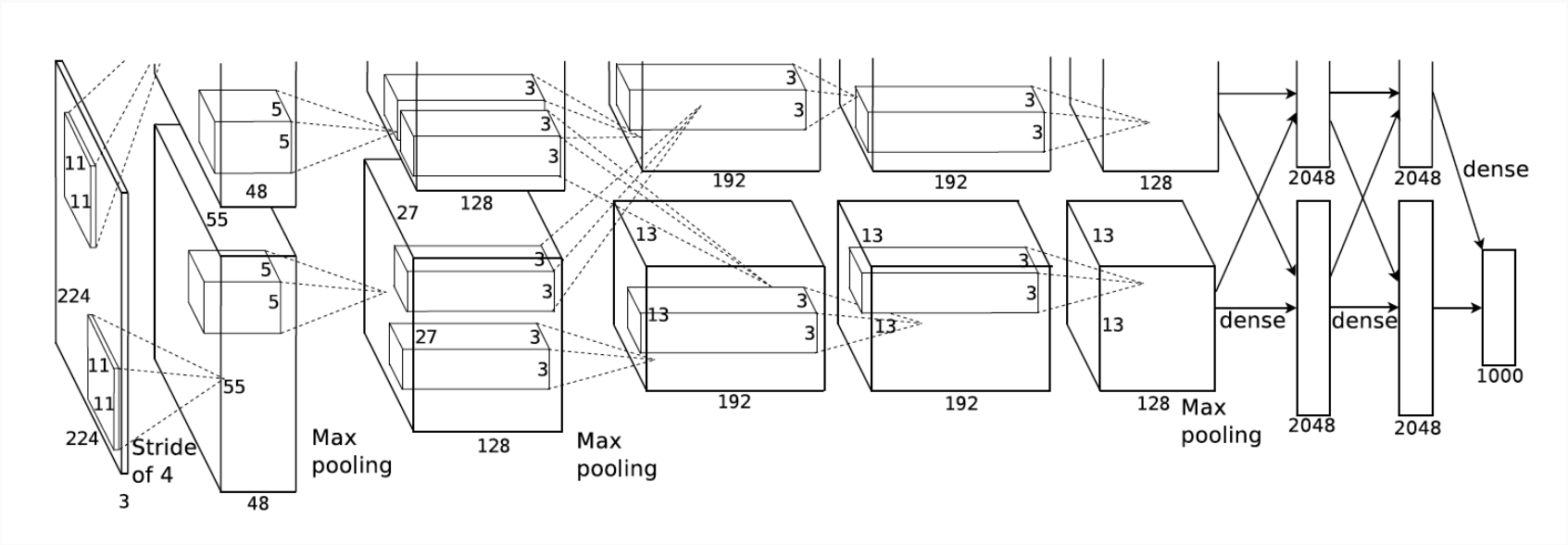
- Классическая архитектура *LeNet*:



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

Imagenet и AlexNet

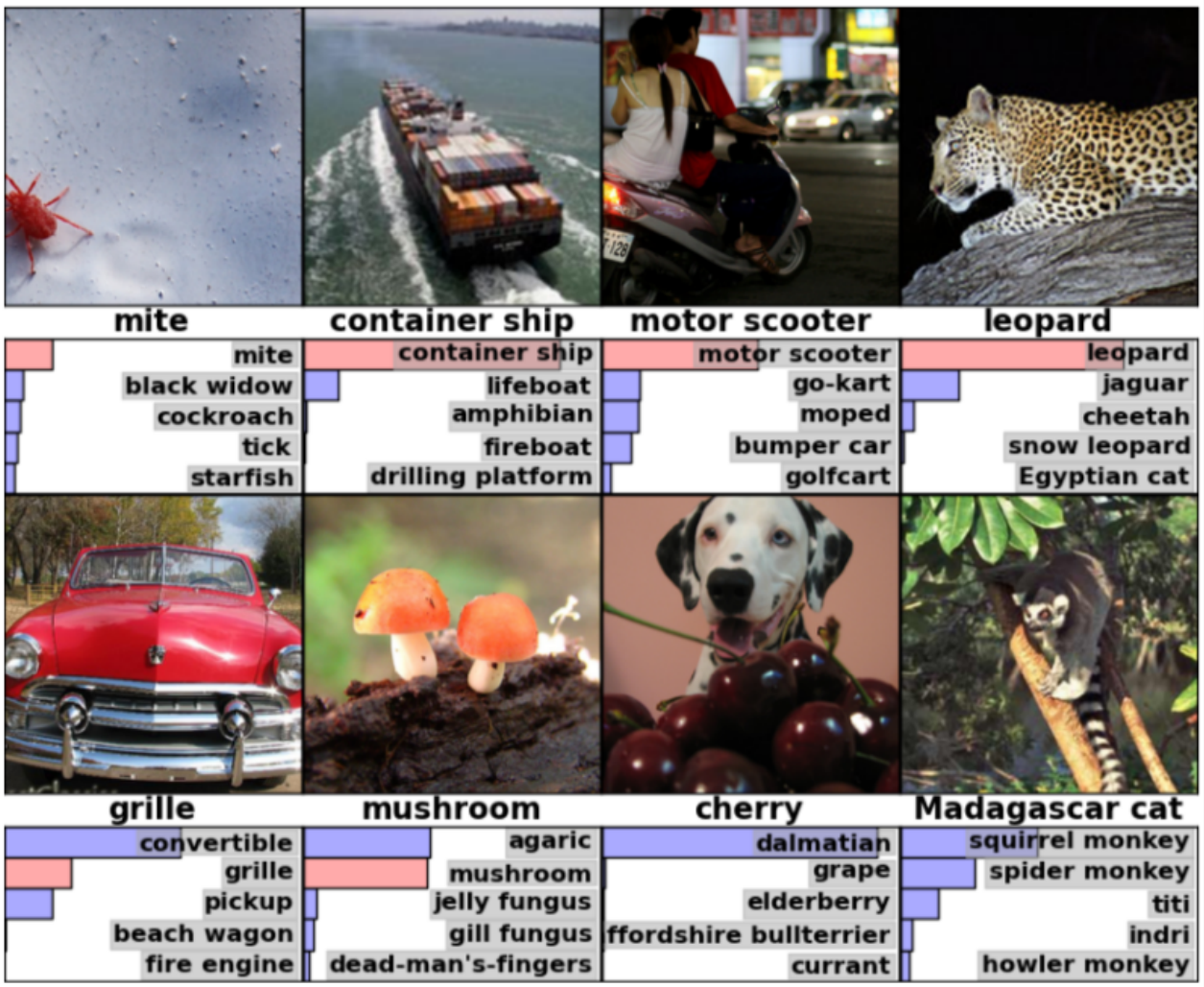
- ImageNet: > 15 миллионов картинок, ≈ 22000 категорий.
- Размечены руками (Amazon Mechanical Turk).
- Классическая сеть AlexNet (Krizhevsky, Sutskever, Hinton, 2012).



Предобработка данных

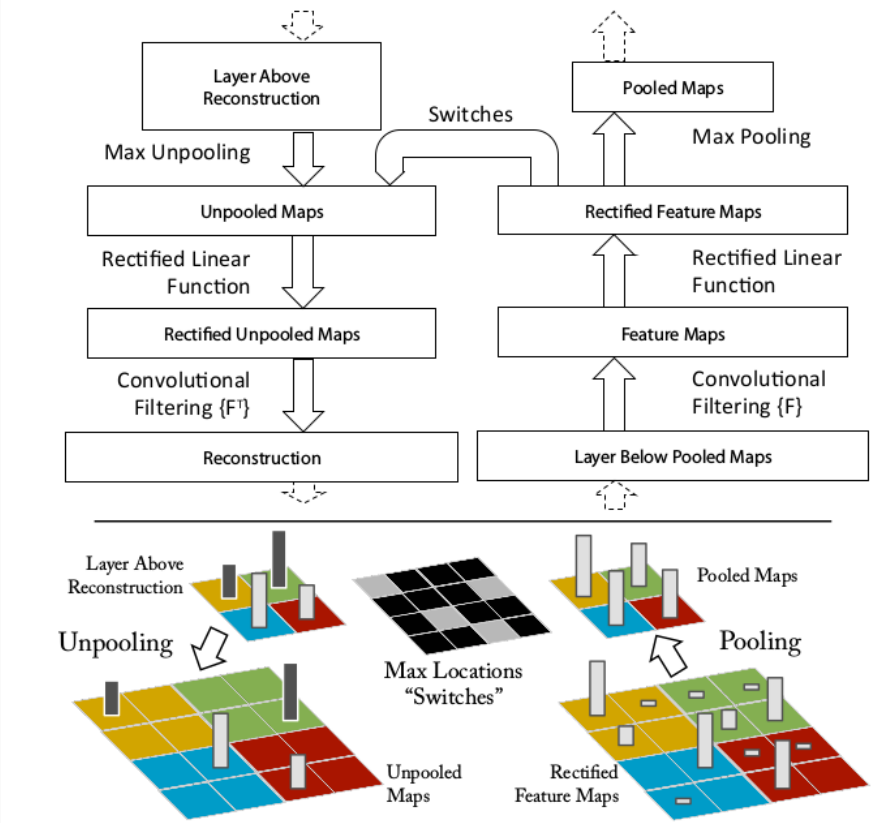
- С картинками можно делать data augmentation: изменять данные без изменения правильного ответа.
- (Krizhevsky, Sutskever, Hinton, 2012):
  - вырежем из 256×256 окна 224×224 (во время применения вырежем пять таких окон и усредним результаты);
  - отразим картинку по горизонтали;
  - поменяем интенсивности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



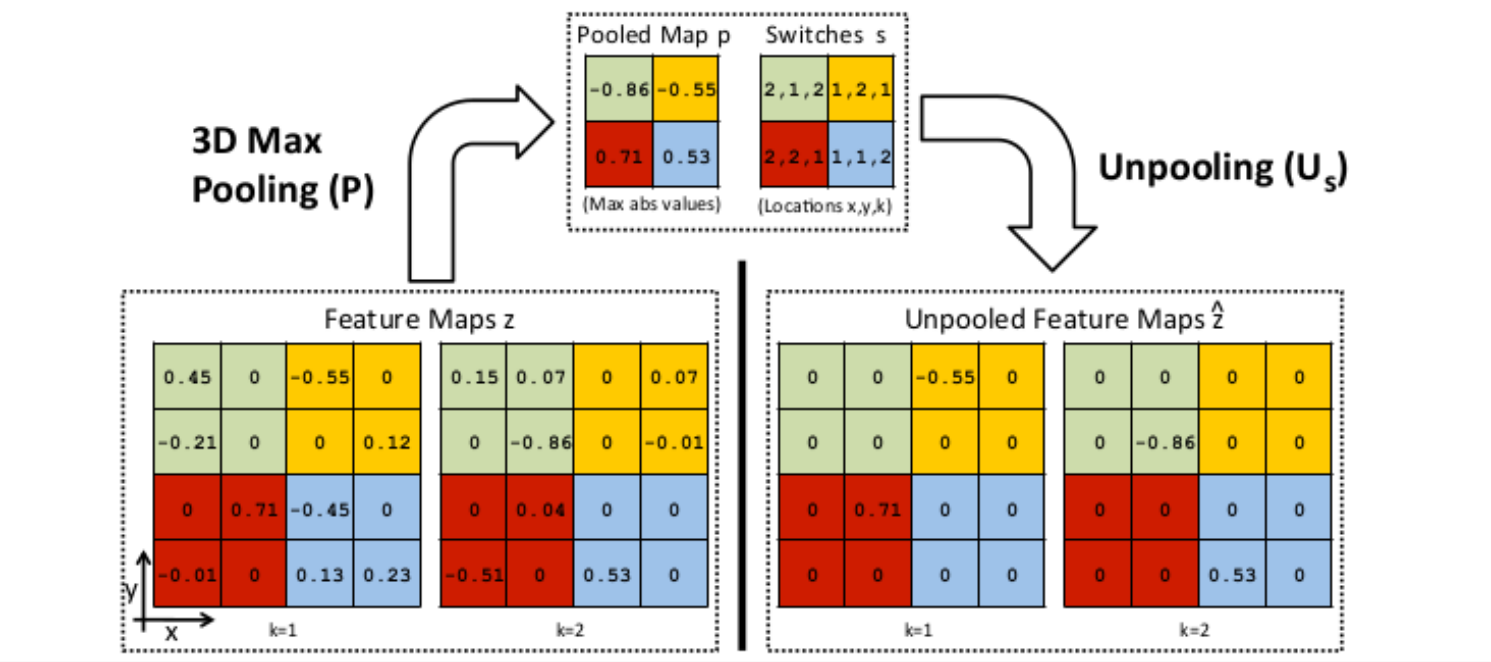
Деконволюция

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



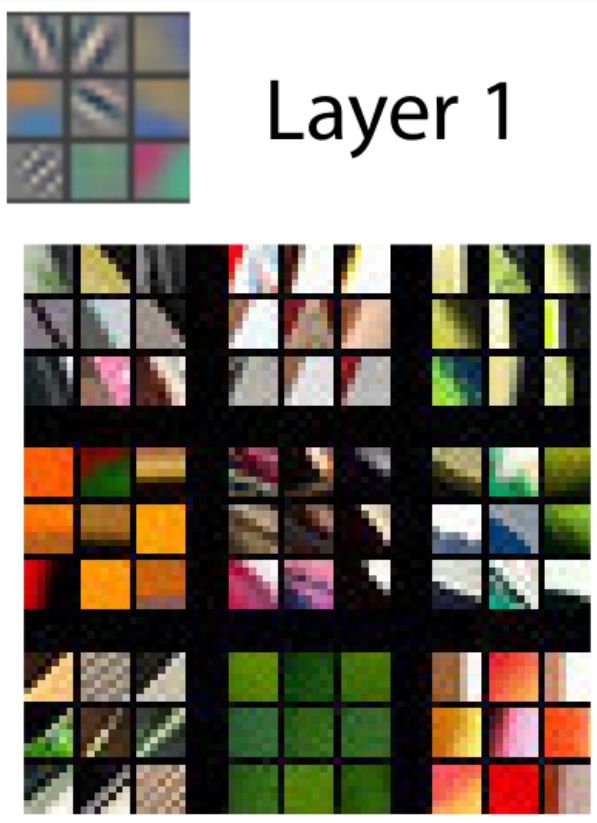
Деконволюция

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

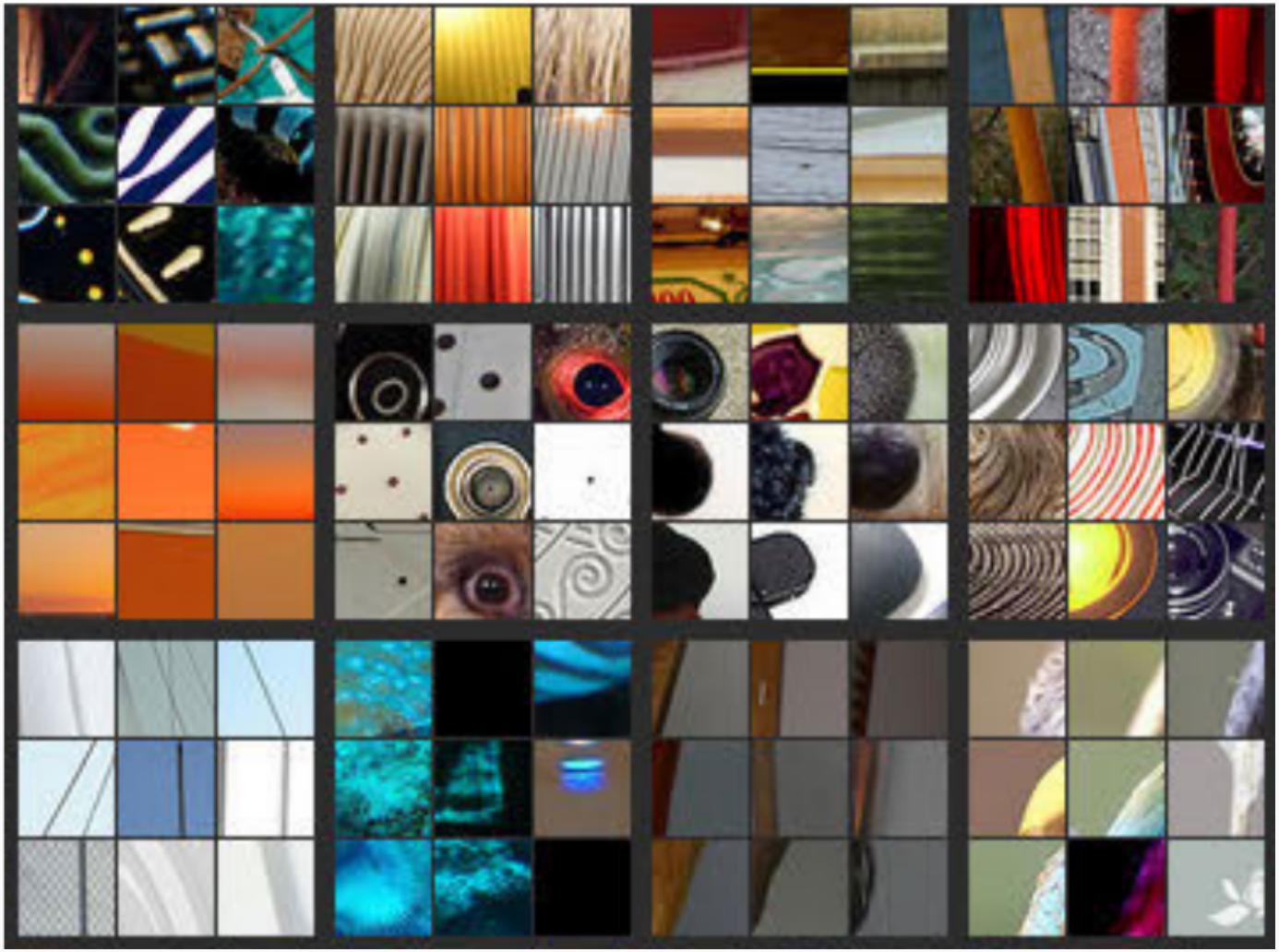


Уровень 1

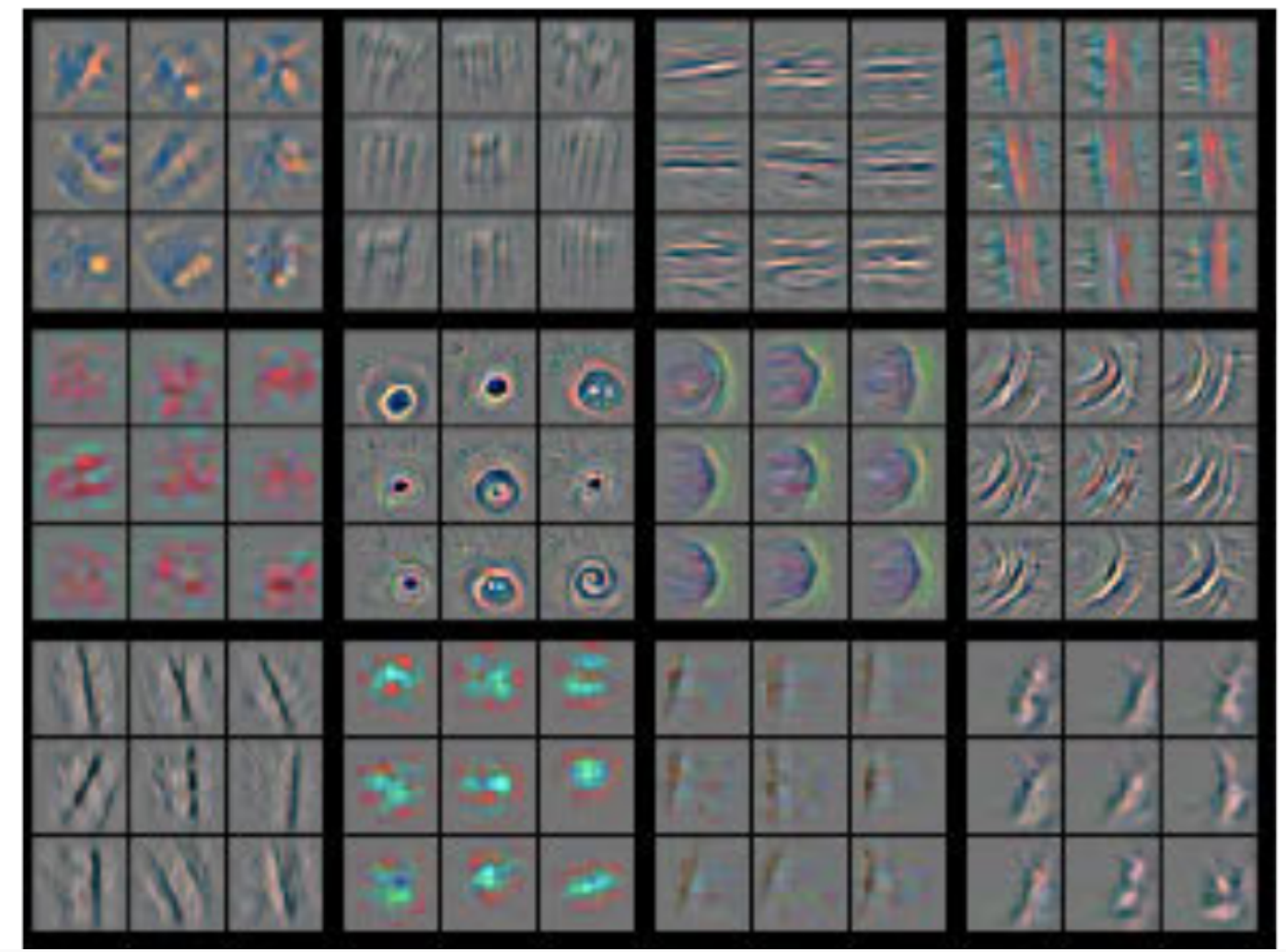
(Zeiler, Fergus, 2014): визуализируем признаки через деконволюцию. Это сеть, обученная на ImageNet.



(Zeiler, Fergus, 2014): уровень 2

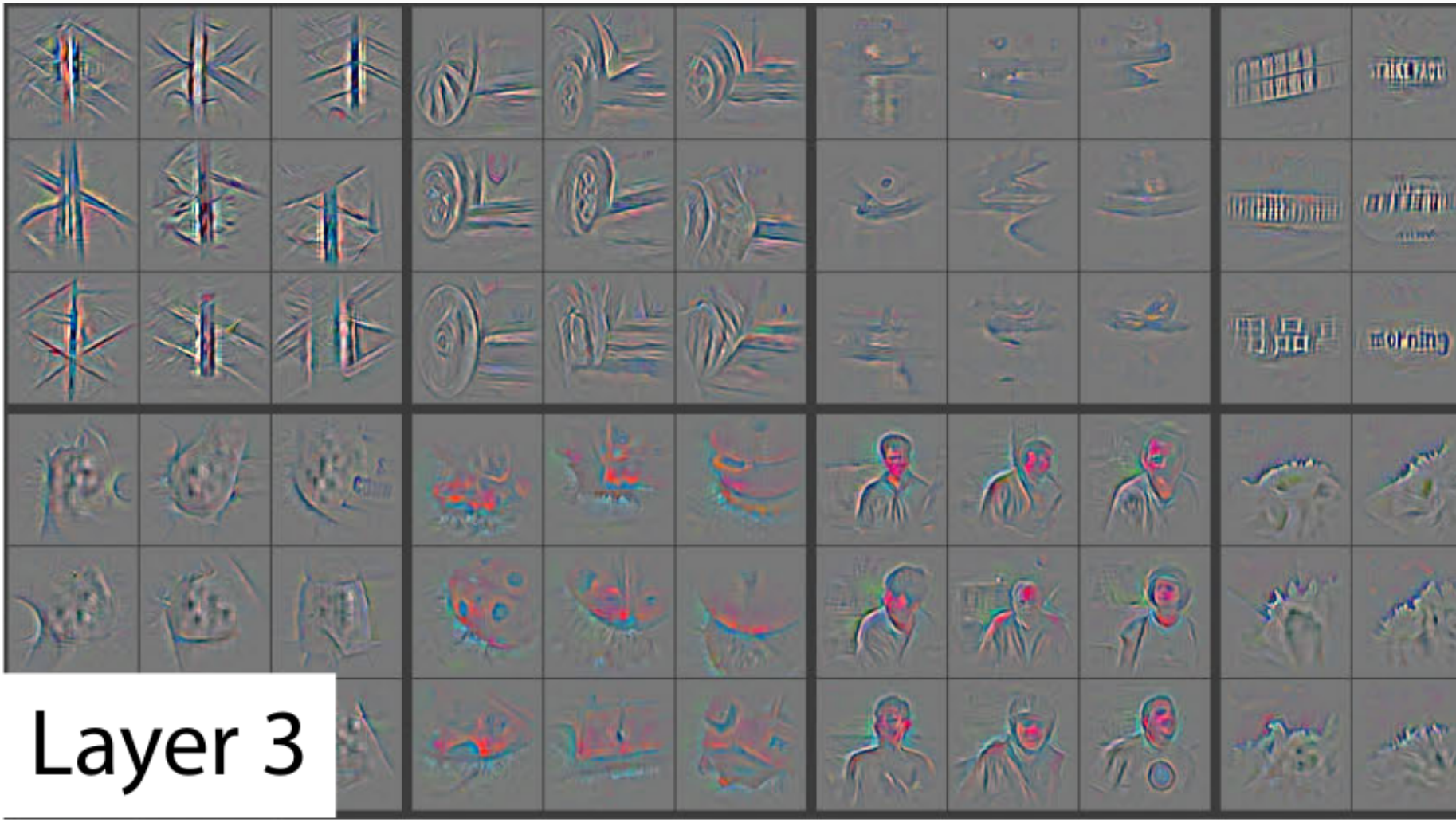


(Zeiler, Fergus, 2014): уровень 2

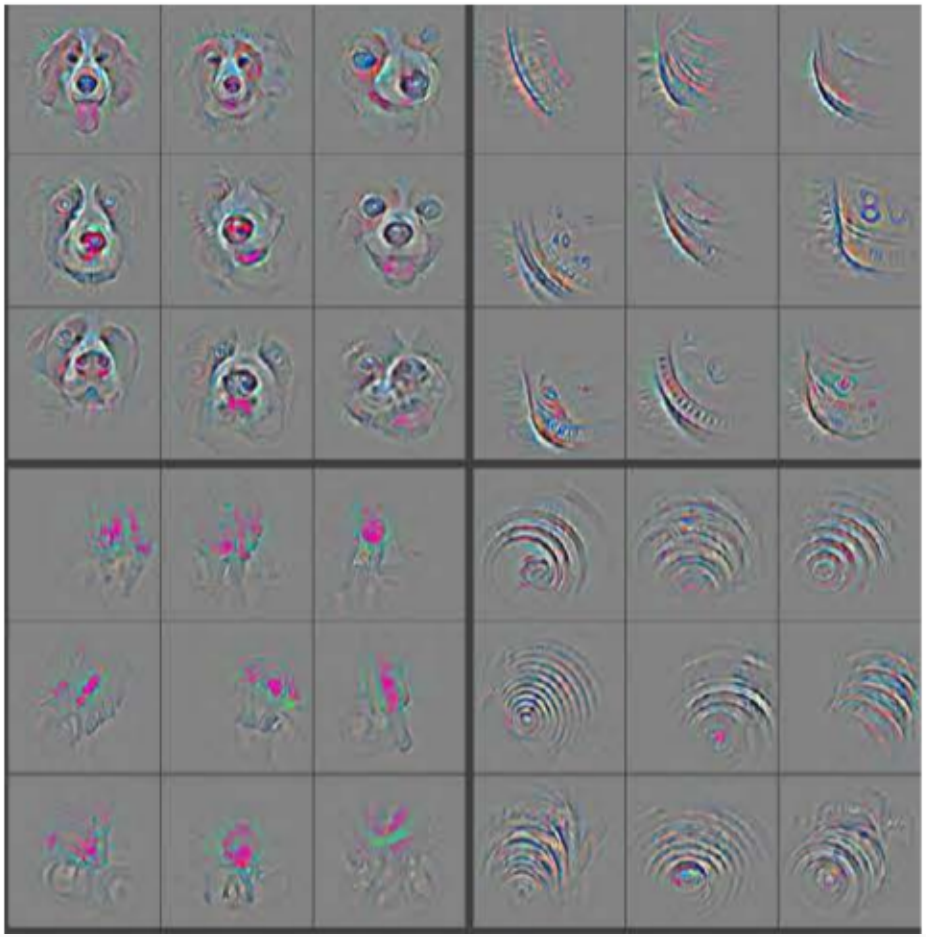


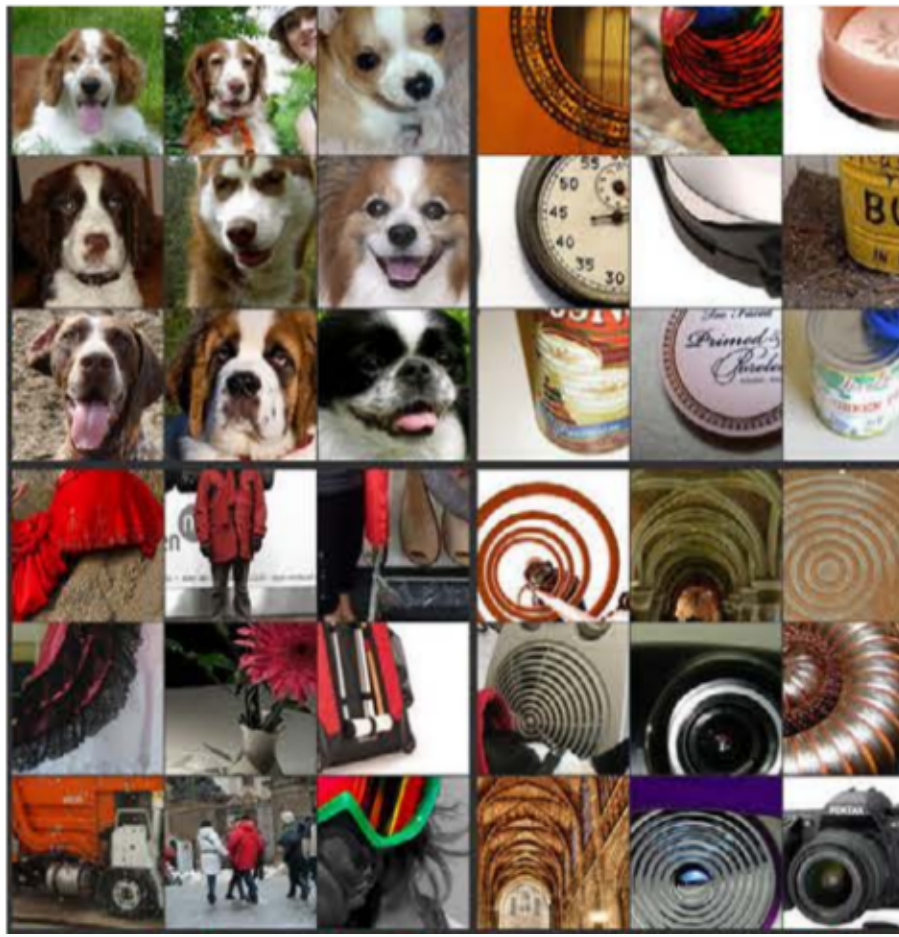
(Zeiler, Fergus, 2014): уровень 3

(Zeiler, Fergus, 2014): уровень 3

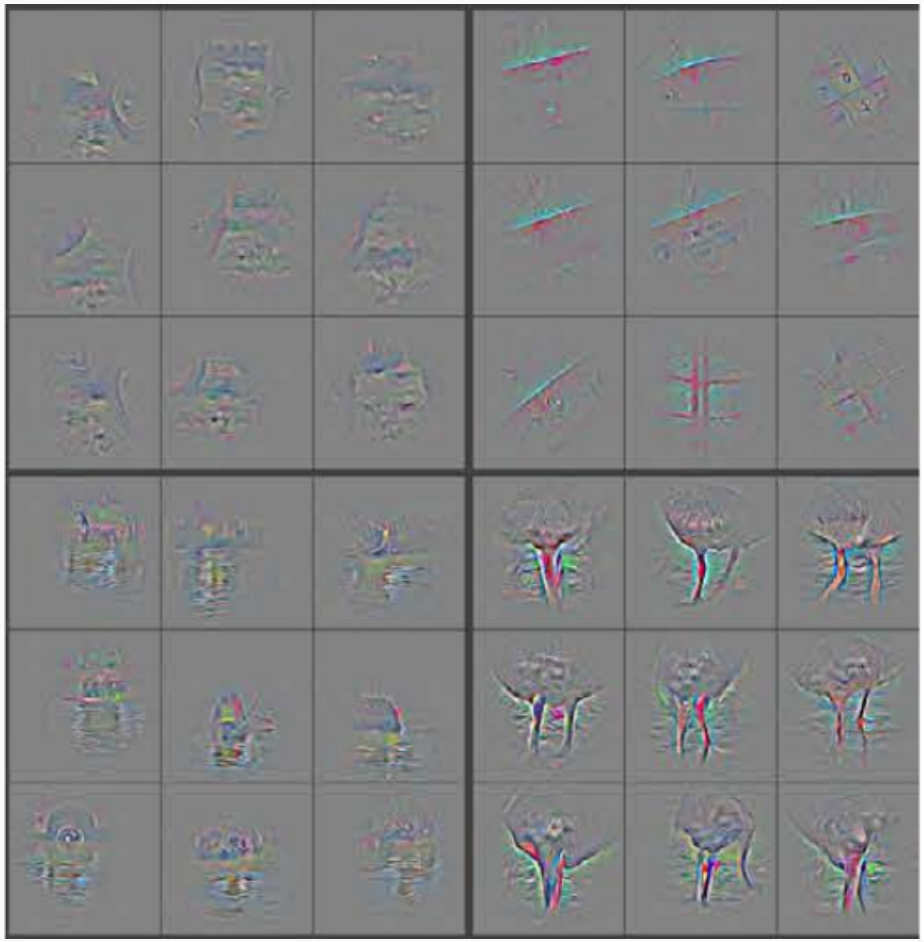


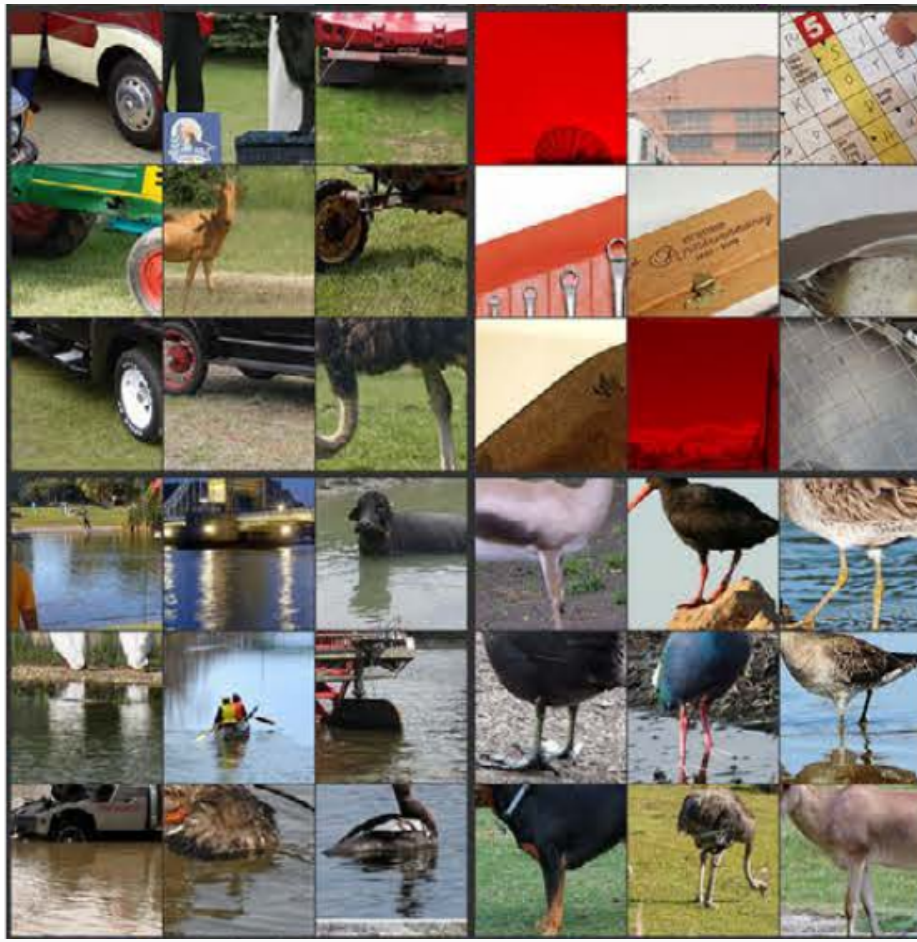
(Zeiler, Fergus, 2014): уровень 4



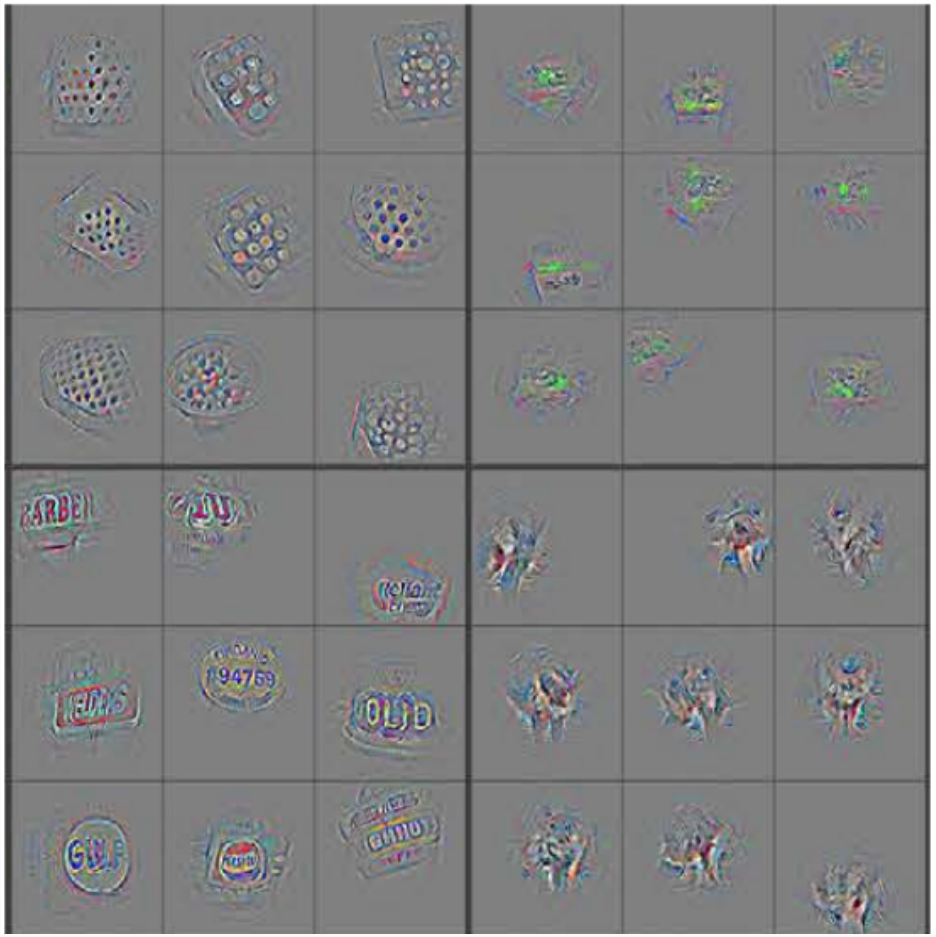


(Zeiler, Fergus, 2014): уровень 4

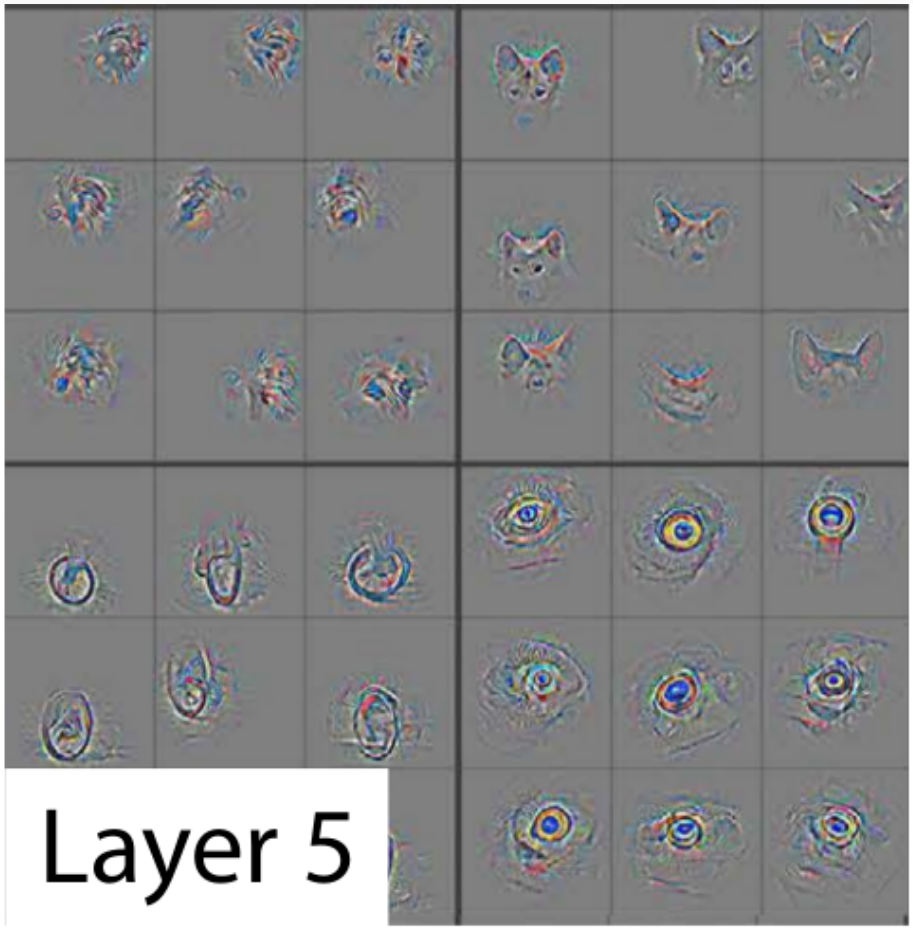




(Zeiler, Fergus, 2014): уровень 5



(Zeiler, Fergus, 2014): уровень 5



Спасибо за внимание!