

Deep learning, Neural Nets for images

Морозов Никита
Романов Даниил



Санкт-Петербург
2023 г.

Definition

Изображение — тензор $M \in \mathbb{R}^{m \times n \times d}$, m — ширина изображения, n — длина. Чаще всего $d = 3$ (3 канала — Red, Green, Blue).

Пусть $X \in \mathbb{R}^{m \times n \times d}$ — случайная величина "изображение".

Задачи:

- Классификация $f : X \rightarrow \{1, \dots, K\}$, K — число классов
- Сегментация $f : X \rightarrow Y$, $Y \in [0, 1]^{m \times n}$,
 $Y_{ij} = P(X_{ij} \in \text{segment})$

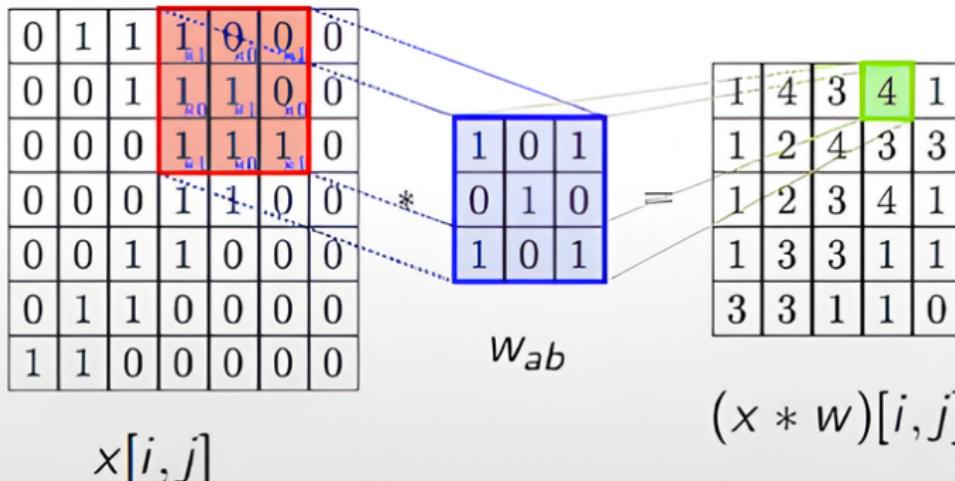
Сверточный слой нейронов

$x[i, j]$ – исходные признаки, пиксели $n \times m$ изображения

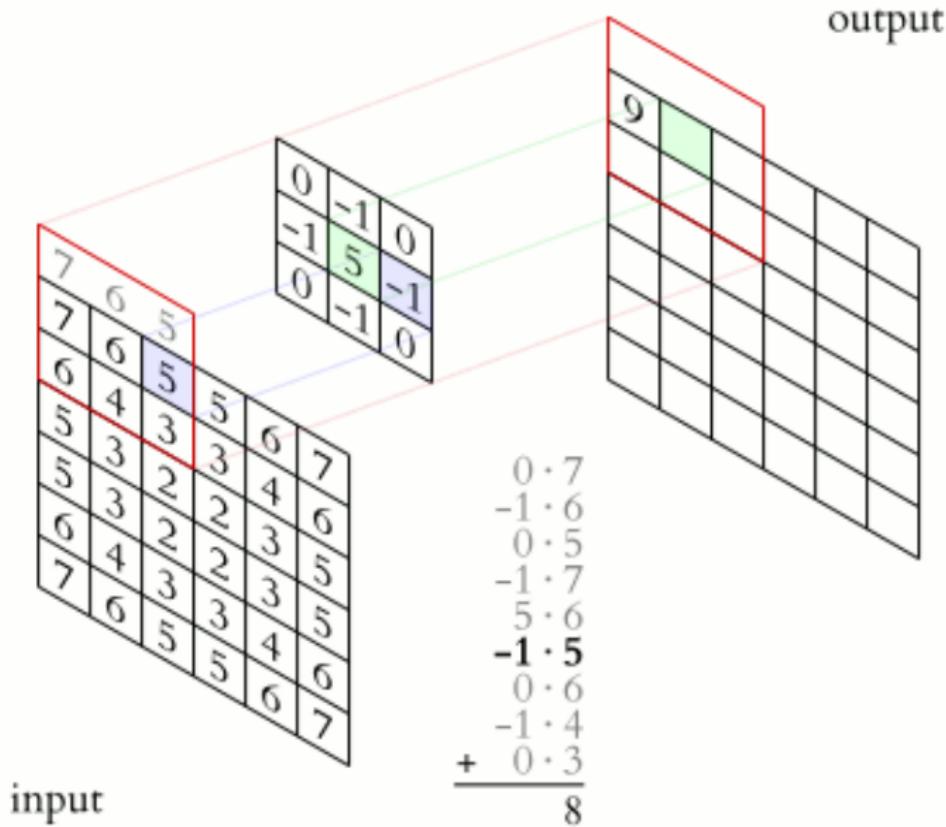
w_{ab} – ядро свертки, $a = -A, \dots, +A$, $b = -B, \dots, +B$

Неполносвязный свёрточный нейрон с $(2A + 1)(2B + 1)$ весами:

$$(x * w)[i, j] = \sum_{a=-A}^A \sum_{b=-B}^B w_{a,b} x[i + a, j + b]$$



Пример свертки



Pooling слой нейронов

Объединяющий нейрон – это необучаемая свертка с шагом $h > 1$, агрегирующая данные прямоугольной области $h \times h$:

$$y[i, j] = F(x[h_i, h_j], \dots, x[h_i + h - 1, h_j + h - 1]),$$

где F – агрегирующая функция: max, average, min.

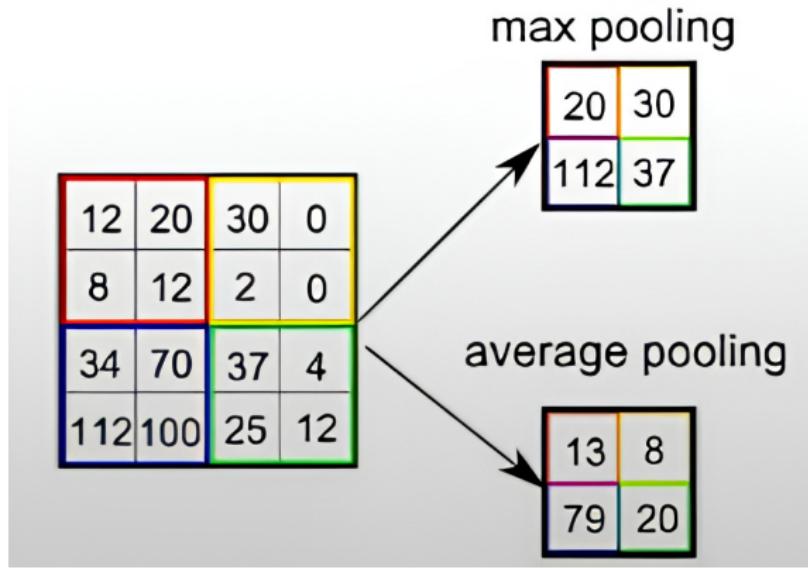
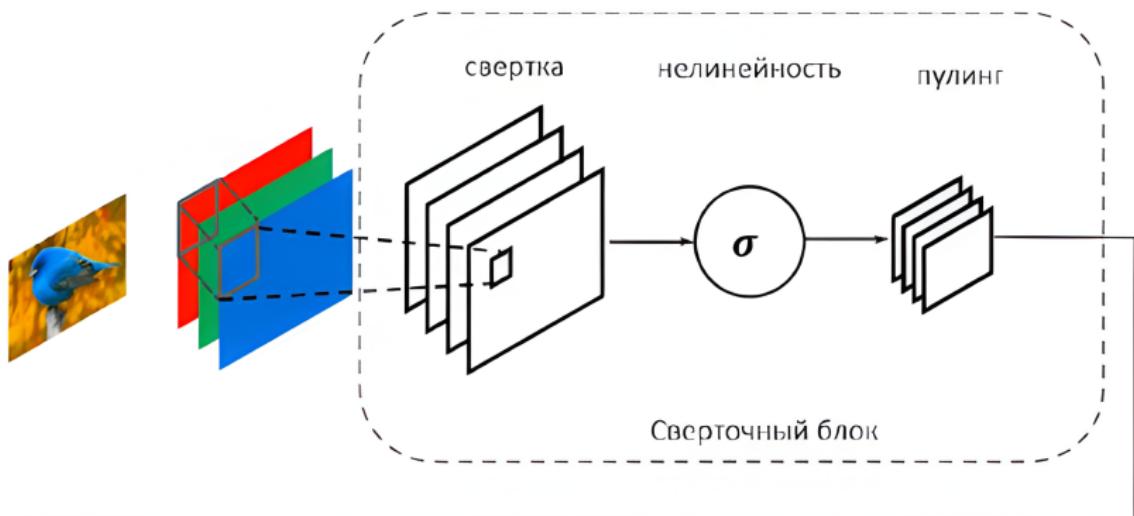
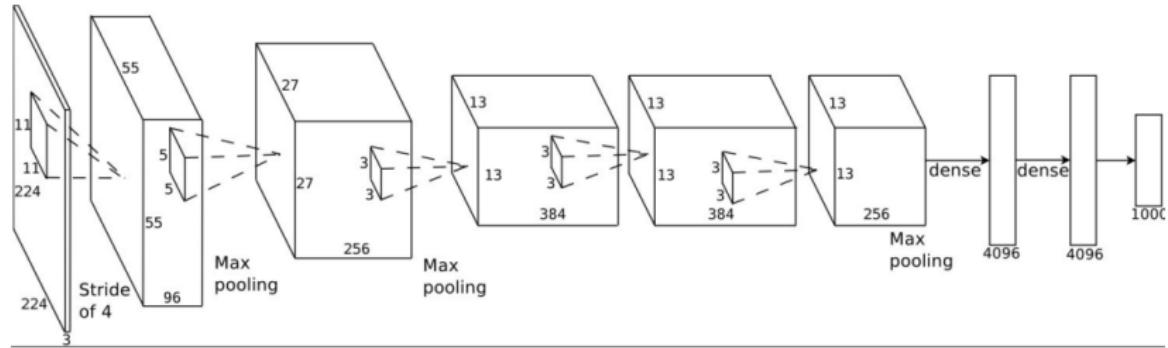


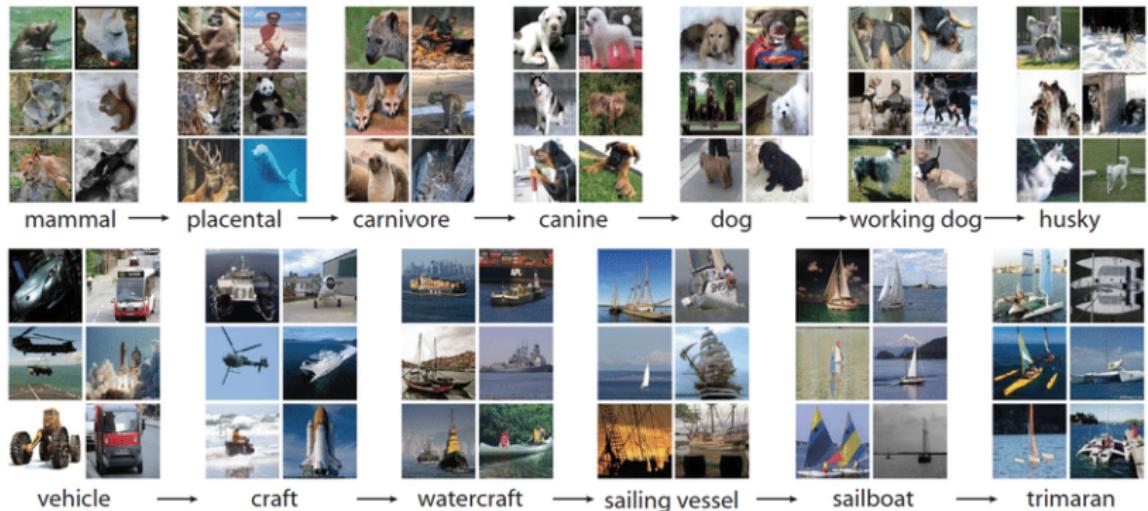
Схема сверточной сети



Уменьшение размерности при помощи CNN



ImageNet



Приемы используемые в CNN

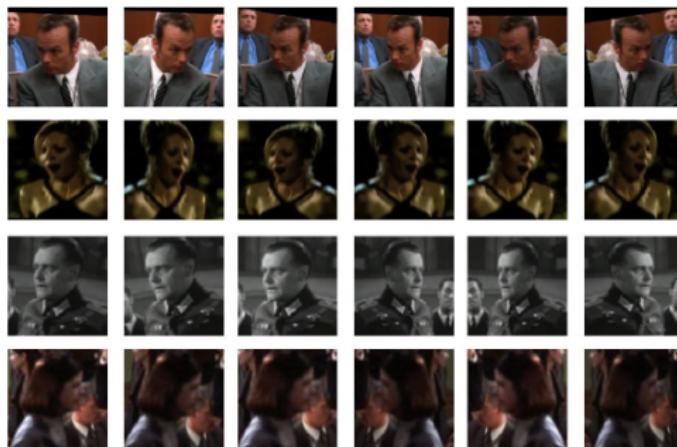
- Подбор числа слоёв и их размерности
- Dropout
- batch normalization
- augmentation(Только для картинок)

Augmentation

Definition

Аугментация (Augmentation) — увеличение объёма тренировочной выборки с помощью различных афинных преобразований изображений: зеркальное отражение, поворот, сдвиг, изменение масштаба.

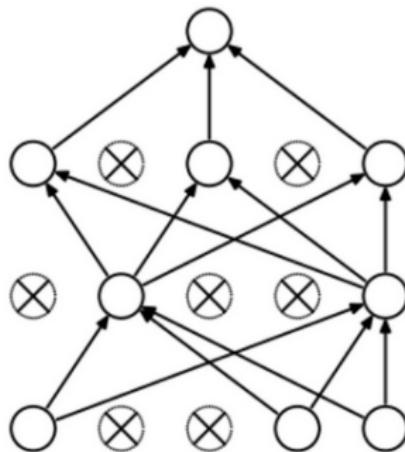
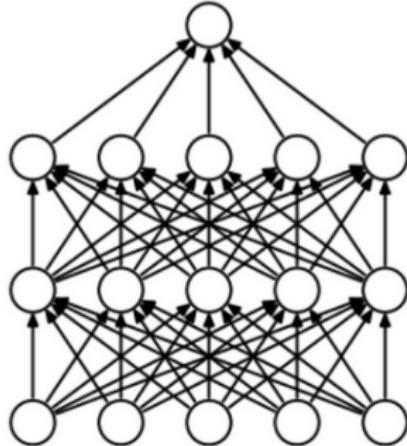
Используется для борьбы с переобучением.



Definition

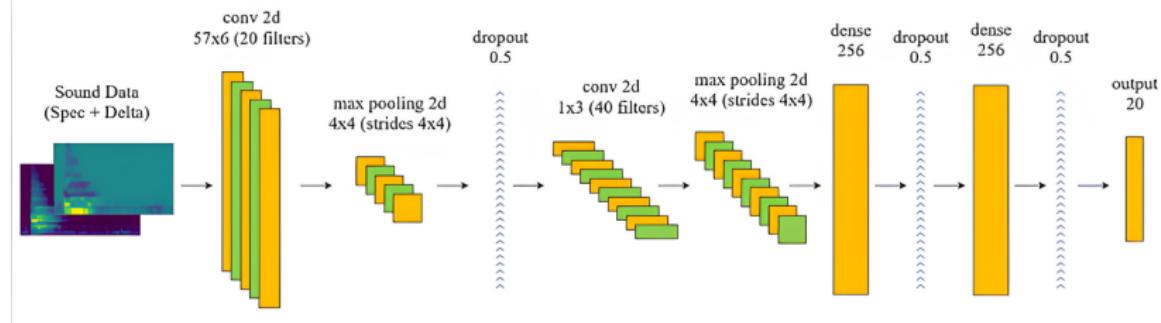
Dropout — отключение (зануление) случайных нейронов во время обучения нейросети. Параметр p — доля отключаемых нейронов. Оставшимся ненулевым нейронам присваиваем вес, равный $\frac{1}{1-p}$.

Цель: борьба с переобучением

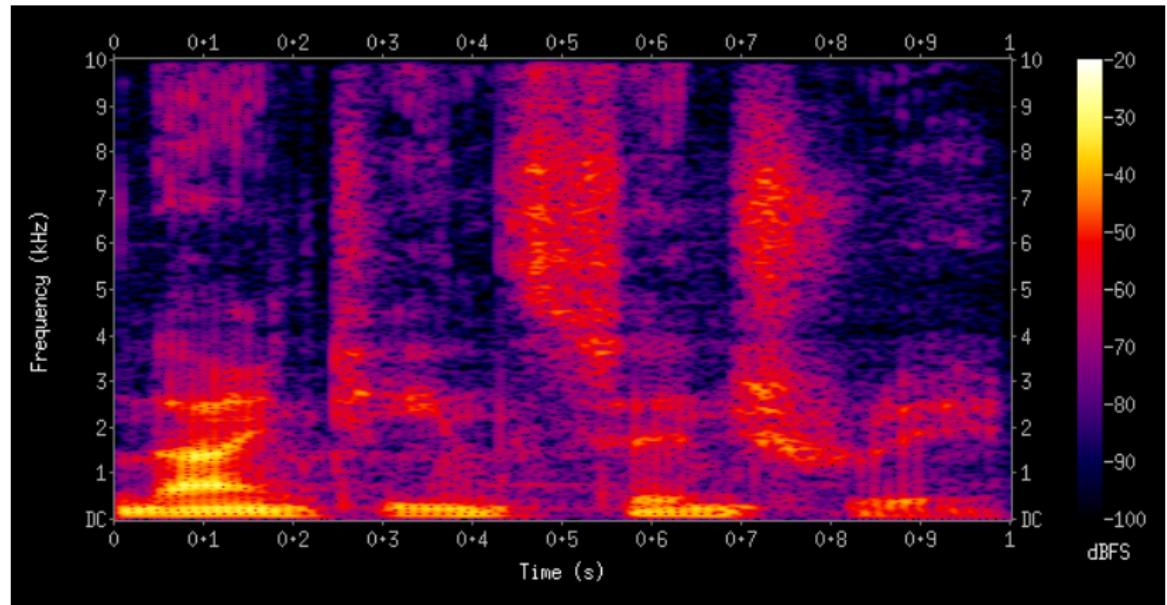


Применение CNN для аудиофайлов

Последовательные фрагменты сигнала представляются векторами спектрального разложения

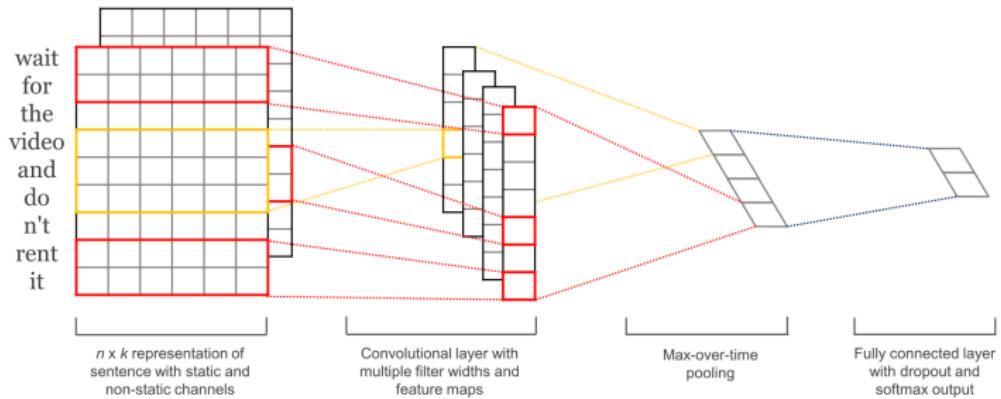


Спектрограмма

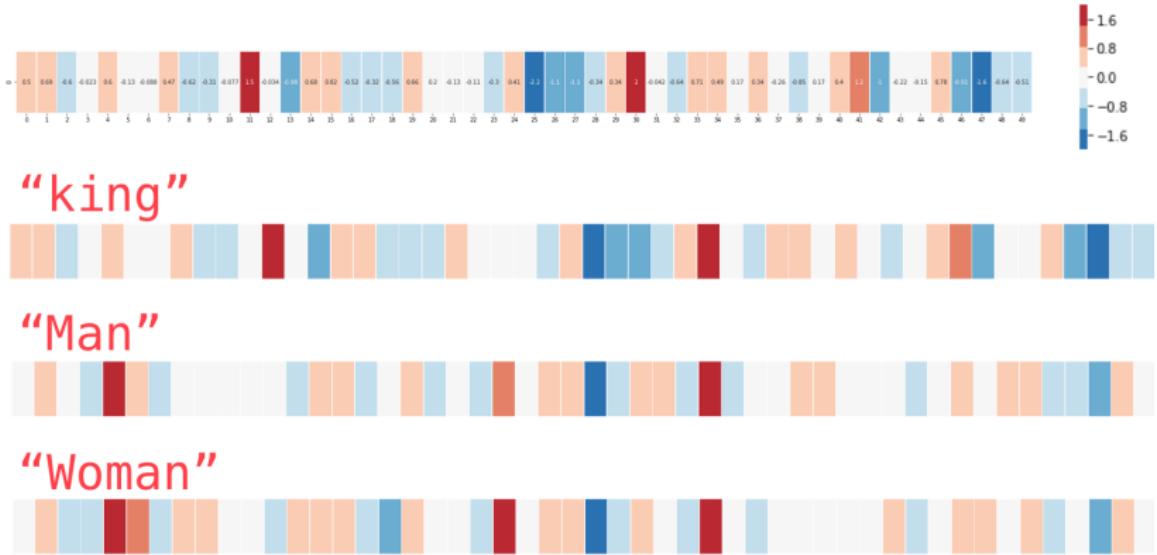


CNN для обработки текста

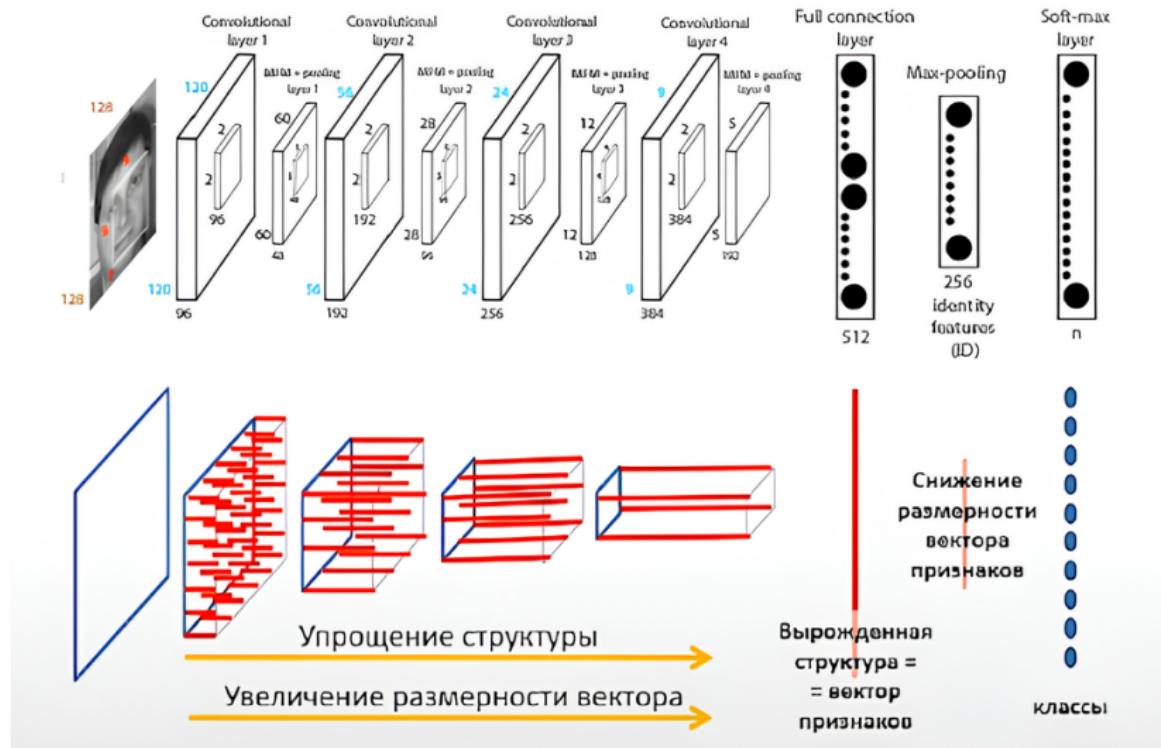
Последовательные слова в тексте представляются векторами с помощью векторных представлений (fastText, word2vec, One-hot encoding и тд.)



word2vec



Обобщение идеи



Проблемы:

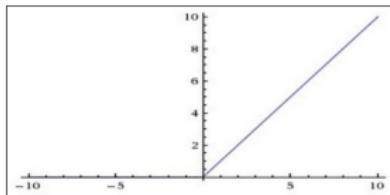
- Необходимость разметки данных для обучения
- Большое количество параметров, следовательно долгое обучение, даже на GPU

Решения:

- Использование размеченных библиотек изображений:
Imagenet (14M изображений, 1000 категорий),
OpenImages (9M изображений, 60K меток, 20K категории)
- Использование предобученной модели (Alexnet, vgg net, Resnet)

Напоминание про функции активации

ReLU (rectified linear unit)



$$f(s) = \max(0, s)$$

$$f'(s) = \begin{cases} 1, & s > 0 \\ \text{rand}(0.01, 0.05), & s \leq 0 \end{cases}$$

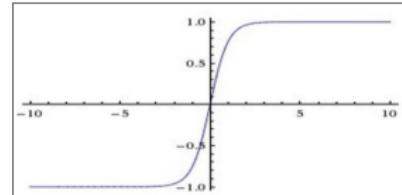
Достоинства

- лишена ресурсоемких операций
- отсекает ненужные детали
- отсутствует разрастание/затухание градиента
- быстрое обучение

Недостатки

- не всегда надежна, в процессе обучения может "умирать"
- сильно зависит от инициализации весов

Гиперболический тангенс



$$f(s) = \frac{e^{2s} - 1}{e^{2s} + 1}$$

$$f'(s) = 1 - f(s)^2$$

Достоинства

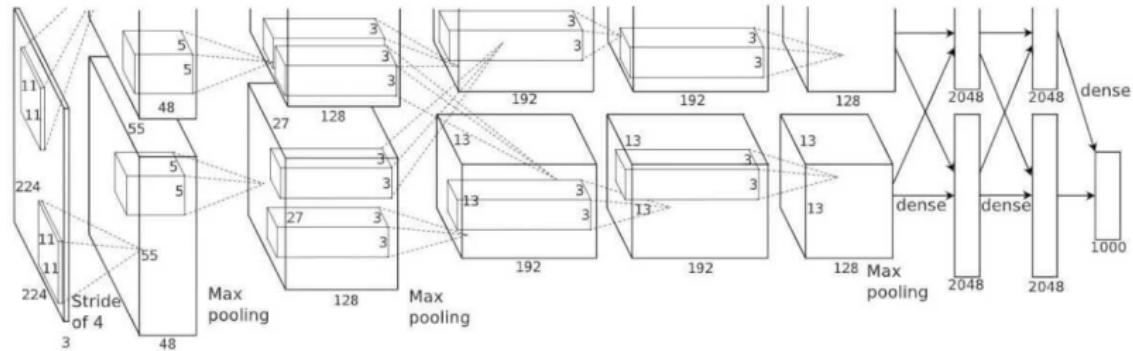
- простое вычисление производной через значение своей функции
- область значений от -1 до 1

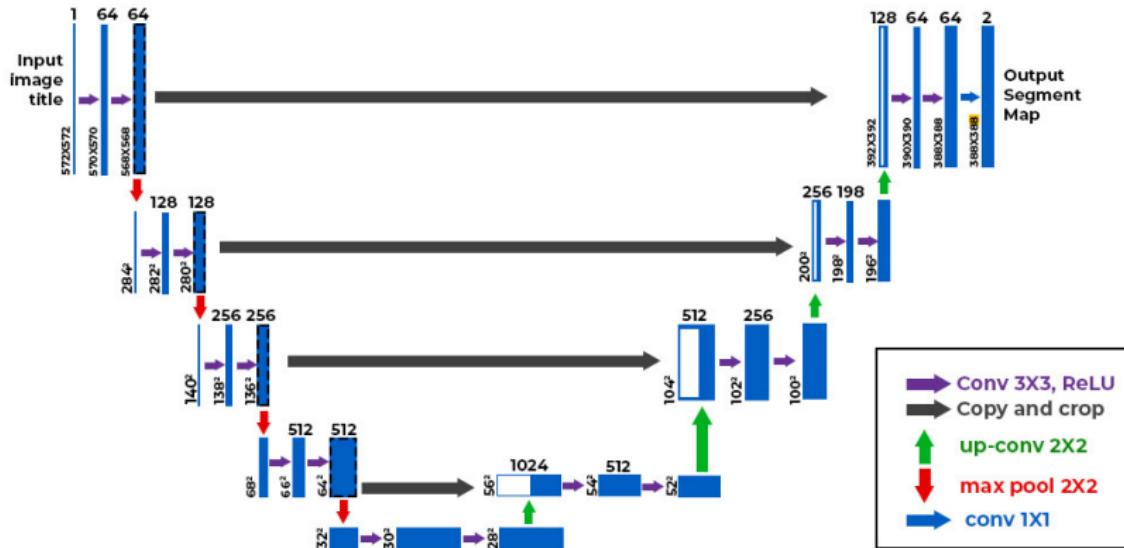
Недостатки

- затухание или увеличение градиента
- ресурсоемкая по сравнению с ReLU

Мертвые нейроны – ситуация, когда входной сигнал ≤ 0 . В результате некоторые нейроны становятся неактивными и не будут участвовать в обработке данных.

AlexNet





vgg16

