

Машинное обучение и анализ данных

Последняя лекция: обзорная

Дьяконов А.Г.

**Московский государственный университет
имени М.В. Ломоносова (Москва, Россия)**

Цель обзора

что осталось за рамками курса
специализированные задачи:
изображения, текст, звук, ...

современные методы решения задач
~ DL

Глубокое обучение (DL) – революция в машинном обучении

Results IJCNN2011 COMPETITION

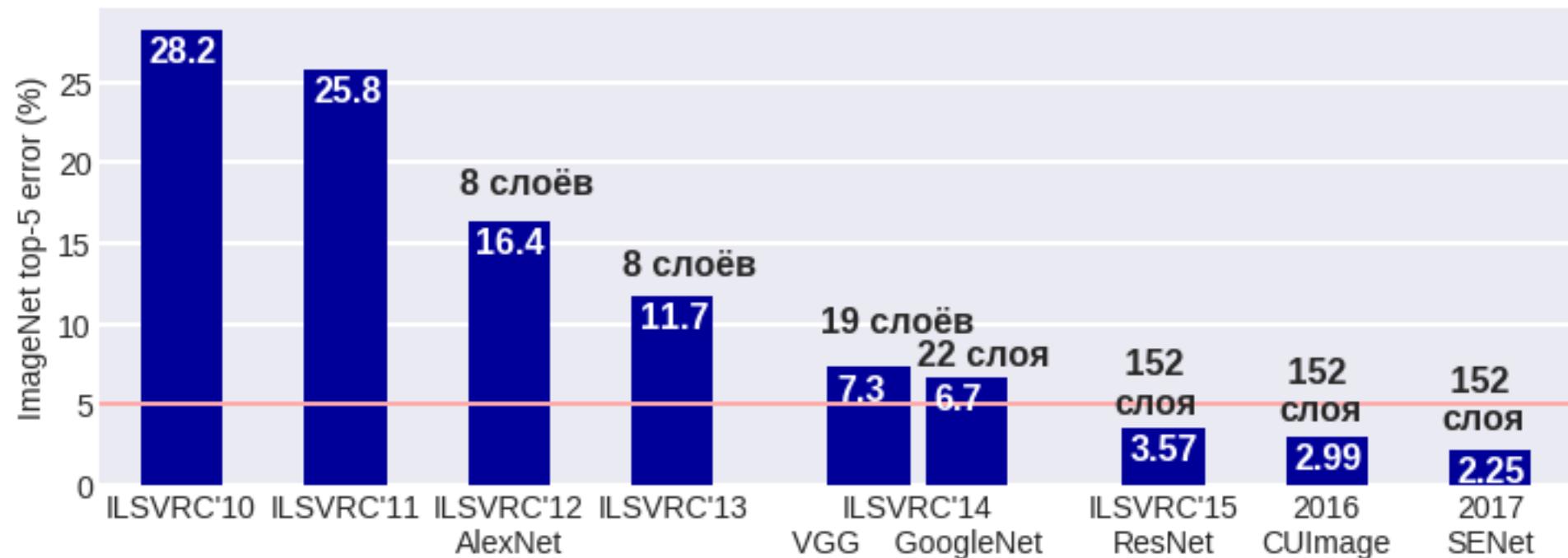
TEAM	METHOD	TOTAL	SUBSET
			ALL SIGNS ▾
[3] IDSIA ★	Committee of CNNs	99.46%	99.46%
[1] INI-RTCV ★	Human Performance	98.84%	98.84%
[4] sermanet ★	Multi-Scale CNNs	98.31%	98.31%
[2] CAOR ★	Random Forests	96.14%	96.14%
[6] INI-RTCV	LDA on HOG 2	95.68%	95.68%
[5] INI-RTCV	LDA on HOG 1	93.18%	93.18%
[7] INI-RTCV	LDA on HOG 3	92.34%	92.34%



2011 год

На соревновании по распознаванию дорожных знаков CNN
превзошёл точность человека!

Глубокое обучение (DL) – революция в машинном обучении



ошибка человека – 5.1

Глубокое обучение (DL) – революция в машинном обучении



Krizhevsky, A., Sutskever, I. and Hinton, G. E., ImageNet Classification with Deep Convolutional Neural Networks, NIPS, 2012

Причины прогресса

- **Интенсивные исследования в ML (+ обмен идеями)**
- **Огромные наборы данных (+ интернет)**
- **Hardware (CPUs, GPUs,)**
- **Новые приёмы в обучении (Dropout, BN и т.п.)**
- **Архитектурные: возможность инвариантности к некоторым преобразованиям (например масштабу)**
- **Интерес бизнеса**

Графический процессор	Год выпуска	Техпроцесс	Гигафлопсов, для 32-разрядных вычислений	Гигафлопсов, для 64-разрядных вычислений
GeForce GTX 1080 (GP104)	2016	16 нм	8873 (Boost)	277 (Boost)
GeForce GTX 780 Ti (GK110-425-B1)	2013	28 нм	5046	210
GeForce GTX 680 (GK104)	2012	28 нм	3090,4	128,8
Radeon HD 7970 (Tahiti)	2012	28 нм	3789	947-1024
Radeon HD 6990 (2xCayman)	2011	40 нм	$2 \times 2550 = 5100$	$2 \times 637 = 1274$

<https://ru.wikipedia.org/wiki/FLOPS>

Данные – причина интенсивного развития ML

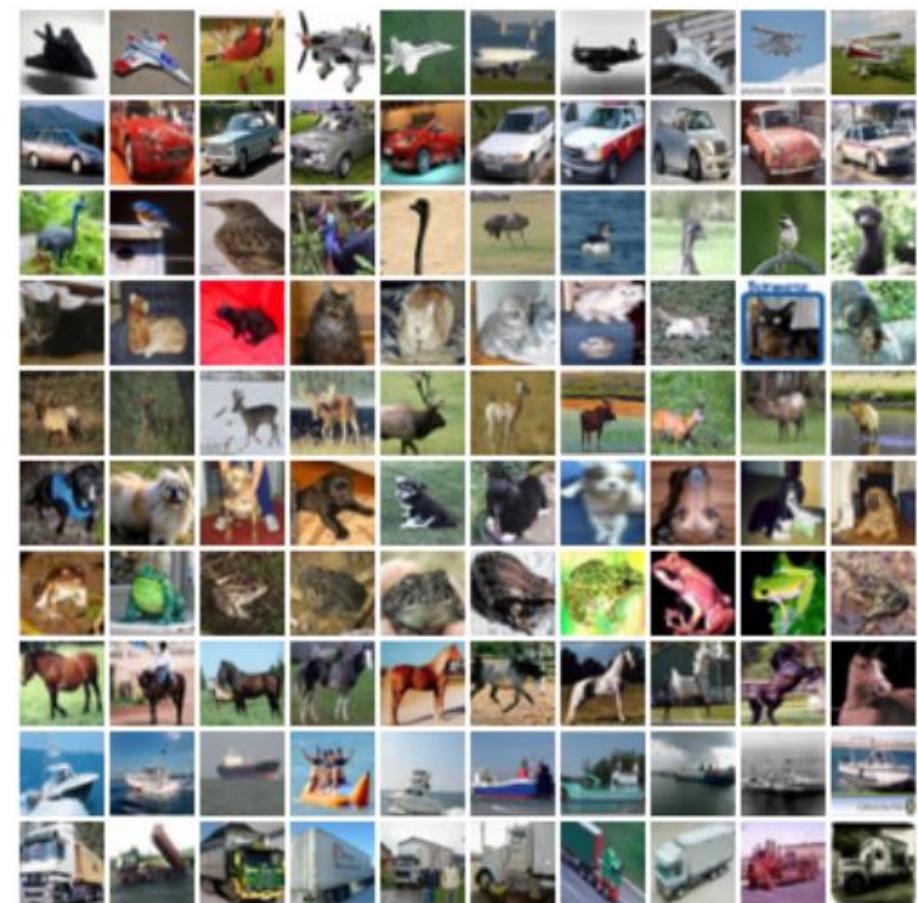
IMAGENET

Класс «Молотки» 1.2 млн. 225x 225-
изображений, 1000 классов



CIFAR10

32x32 цветные, 50k – обучение,
10k – тест



Данные – причина интенсивного развития ML



<http://places.csail.mit.edu/>

MNIST

28×28 изображений (оттенки серого), 60k обучение, 10k тест

```
1 1 8 3 6 1 0 3 1 0 0 1 1 2 7 3 0 4 6 5  
2 6 4 7 1 8 9 9 3 0 7 1 0 2 0 3 5 4 6 5  
8 6 3 7 5 8 0 9 1 0 3 1 2 2 3 3 6 4 7 5  
0 6 2 7 9 8 5 9 2 1 1 4 4 5 6 4 1 2 5 3  
9 3 9 0 5 9 6 5 7 4 1 3 4 0 4 8 0 4 3 6  
8 7 6 0 9 7 5 7 2 1 1 6 8 9 4 1 5 2 2 9  
0 3 9 6 7 2 0 3 5 4 3 6 5 8 9 5 4 7 4 2  
1 3 4 8 9 1 9 2 8 7 9 1 8 7 4 1 3 1 1 0  
2 3 9 4 9 2 1 6 8 4 1 7 4 4 9 2 8 7 2 4  
4 2 1 9 7 2 8 7 6 9 2 2 3 8 1 6 5 1 1 0  
4 0 9 1 1 2 4 3 2 7 3 8 6 9 0 5 6 0 7 6  
2 6 4 5 8 3 1 5 1 9 2 7 4 4 8 1 5 8 9  
5 6 7 9 9 3 7 0 9 0 6 6 2 3 9 0 1 5 4 8  
0 9 4 1 2 8 7 1 2 6 1 0 3 0 1 1 8 2 0 3  
9 4 0 5 0 6 1 7 7 8 1 9 1 0 5 1 2 2 7 3  
5 4 4 7 1 8 3 9 6 0 3 1 1 2 6 3 5 7 6 8  
3 9 5 8 5 7 4 1 1 3 1 7 5 5 5 2 5 8 7 0  
9 7 7 5 0 9 0 0 8 9 2 4 8 1 6 1 6 5 1 8  
3 4 0 5 5 8 3 6 2 3 9 2 1 1 5 2 1 3 2 8  
7 3 7 2 4 6 9 7 2 4 2 8 1 1 3 8 4 0 6 5
```

Cityscapes Dataset

<https://www.cityscapes-dataset.com/>



Данные – причина интенсивного развития ML

QuickDraw



<https://quickdraw.withgoogle.com/>

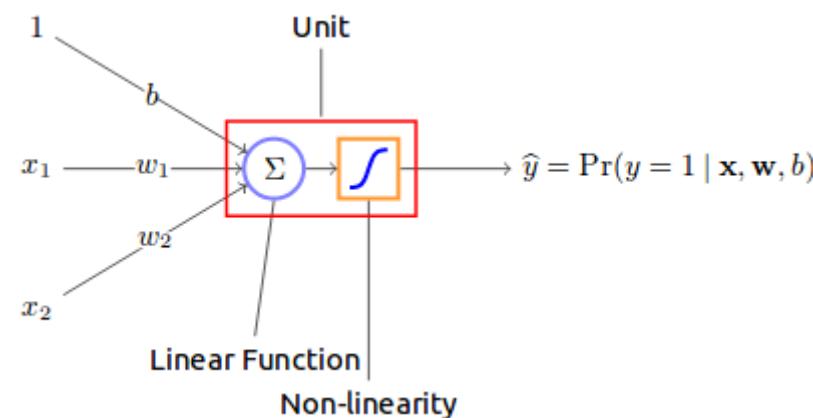
<https://arxiv.org/pdf/1704.03477.pdf>

Наличие библиотек –MNIST на Pytorch:

```
model = nn.Sequential(  
    nn.Conv2d(1, 32, 5), nn.MaxPool2d(3), nn.ReLU(),  
    nn.Conv2d(32, 64, 5), nn.MaxPool2d(2), nn.ReLU(),  
    Flattener(),  
    nn.Linear(256, 200), nn.ReLU(),  
    nn.Linear(200, 10)  
)  
  
nb_epochs, batch_size = 10, 100  
criterion = nn.CrossEntropyLoss()  
optimizer = torch.optim.SGD(model.parameters(), lr = 0.1)  
  
model.to(device)  
criterion.to(device)  
train_input, train_target = train_input.to(device), train_target.to(device)  
  
mu, std = train_input.mean(), train_input.std()  
train_input.sub_(mu).div_(std)  
  
for e in range(nb_epochs):  
    for input, target in zip(train_input.split(batch_size), train_target.split(batch_size)):  
        output = model(input)  
        loss = criterion(output, target)  
        optimizer.zero_grad()  
        loss.backward()  
        optimizer.step()
```

~7s (GTX1080), ~1% test error

Простейшая нейросеть – 1 нейрон



Разделяющая поверхность – линейная

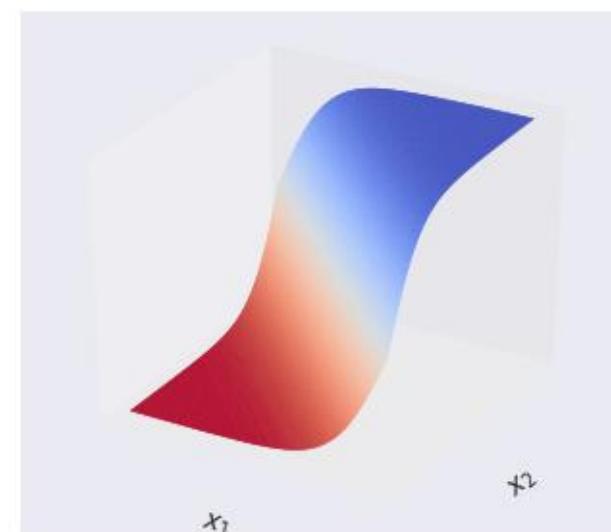
$$a(x) = b + w_1x_1 + \dots + w_nx_n = \sum_{t=0}^n w_t x_t$$

$$h(x) = \sigma(a(x))$$

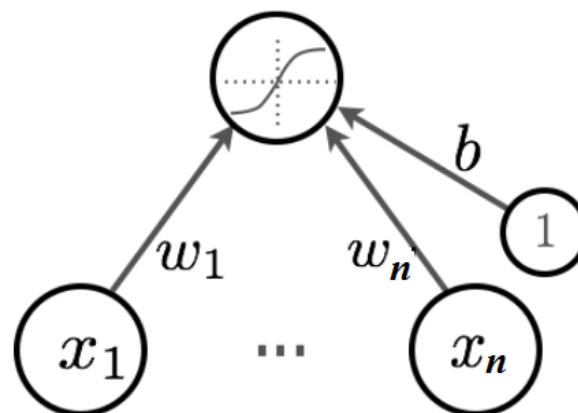
b – смещение

σ – функция активации

w_t – веса связей



Линейные модели – нейросети!



Линейная регрессия

$$a(x) = b + w_1x_1 + \dots + w_nx_n$$

Логистическая регрессия

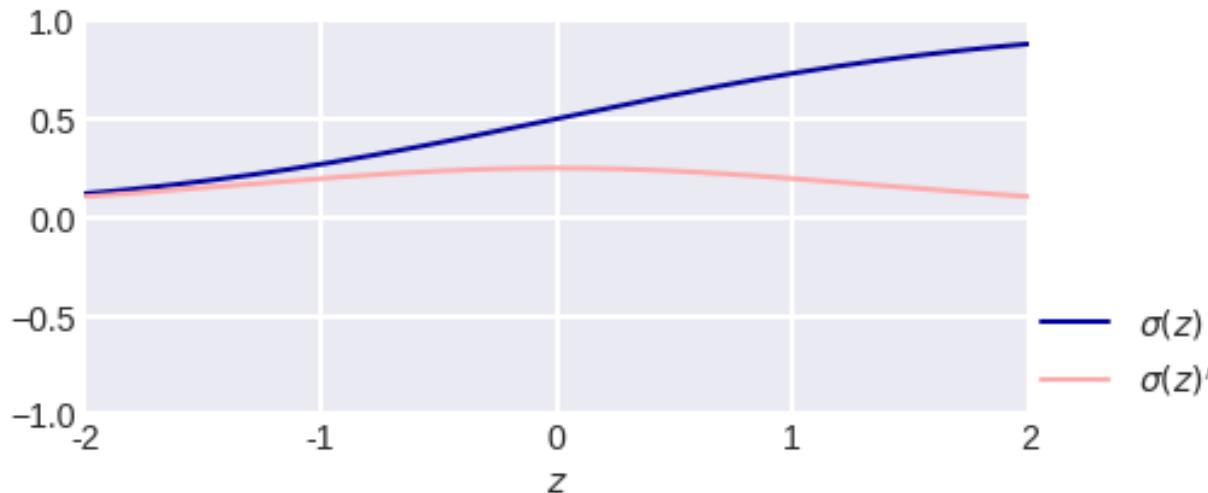
$$a(x) = \sigma(b + w_1x_1 + \dots + w_nx_n)$$

Линейный классификатор

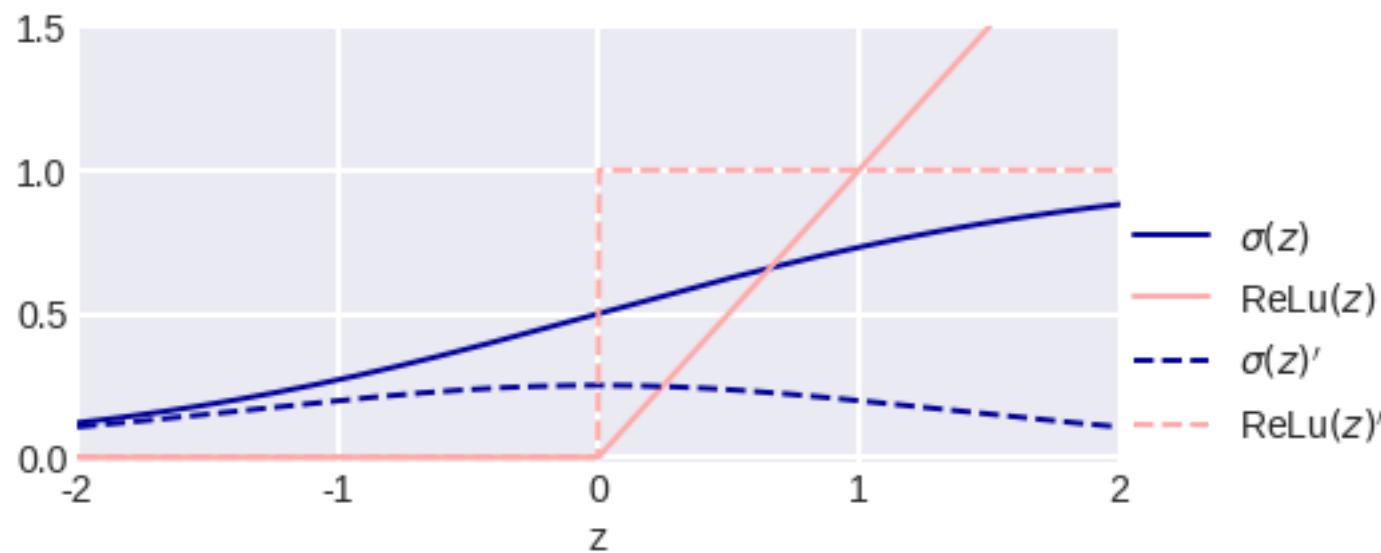
$$a(x) = \text{th}(b + w_1x_1 + \dots + w_nx_n)$$

Функции активации

Сигмоида (sigmoid activation function)



$$\sigma(z) = \frac{1}{1 + e^{-z}} \in (0, 1)$$



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

Функции активации в задачах классификации

$$\text{softmax}(z_1, \dots, z_k) = \frac{1}{\sum_{t=1}^k \exp(z_t)} (\exp(z_1), \dots, \exp(z_k))^T$$

сумма выходов = 1
выходы интерпретируются как вероятности

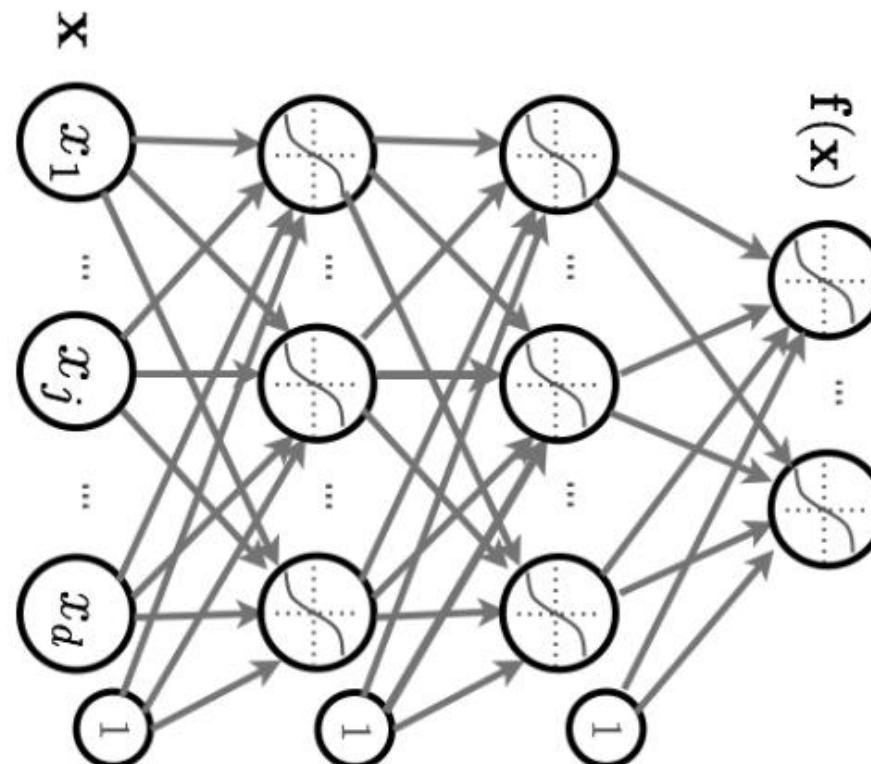
$$[0.5, 0.5, 0.1, 0.7] \rightarrow [0.257, 0.257, 0.172, \color{red}{0.314}]$$

$$[-1.0, 0, 1.0, 0, -1.0] \rightarrow [0.07, 0.18, \color{red}{0.5}, 0.18, 0.07]$$

$$[1.0, 1.0, 1.0, 2.0, 1.0] \rightarrow [0.15, 0.15, 0.15, \color{red}{0.4}, 0.15]$$

Сеть прямого распространения – Feedforward Neural Network (т.е. нет циклов)

все нейроны предыдущего слоя связаны с нейронами следующего



входной слой

один или несколько скрытых слоёв

выходной слой

Обучение

Как принято...
минимизация регуляризованного эмпирического риска

$$\frac{1}{m} \sum_{t=1}^m L(a(x_i | w), y_i) + \lambda R(w) \rightarrow \min_w$$

Задача оптимизации невыпуклая!

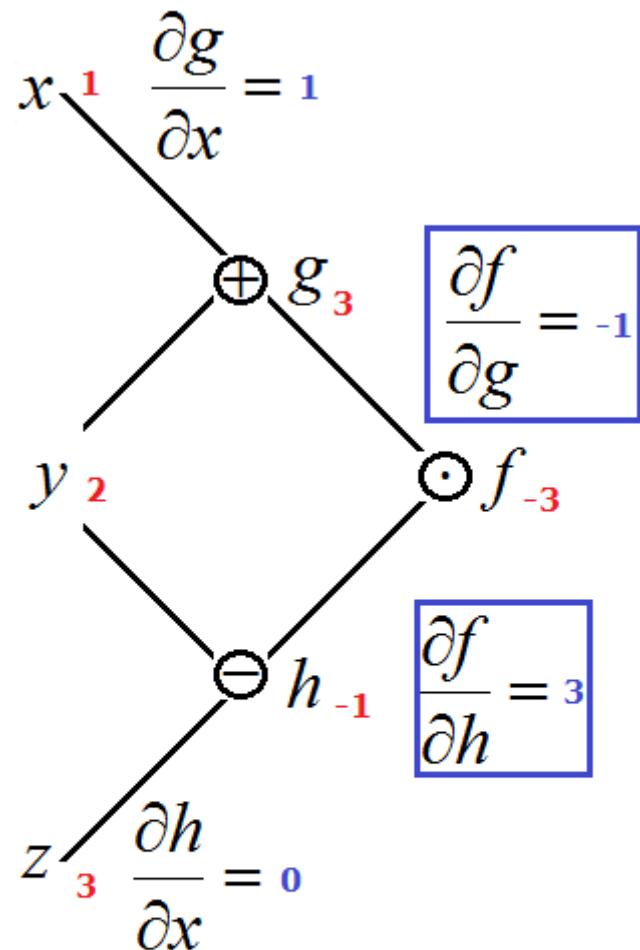
«Настройка» нейронной сети – получение весов w

Метод стохастического градиента

$$w^{(t+1)} = w^{(t)} - \eta \nabla [L(a(x_i | w^{(t)}), y_i) + \lambda R(w^{(t)})]$$

т.к. очень много слагаемых... и так быстрее;)

Вычисление градиента



$$f(x, y, z) = \underbrace{(x + y)}_{g(x, y)} \cdot \underbrace{(y - z)}_{h(y, z)}$$

Как проводится вычисление производных?

«Обратный ход»

$$\frac{\partial f}{\partial x} = \boxed{\frac{\partial f}{\partial g}} \frac{\partial g}{\partial x} + \boxed{\frac{\partial f}{\partial h}} \frac{\partial h}{\partial x}$$

Борьба с переобучением

- Нормировки (**Normalization of Data**)
- Инициализация весов
- Верификация – ранний останов (**Early Stopping**)
- Настройка темпа обучения (**Learning Rate**)
- Мини-батчи (**Mini-Batches**) / Batch-обучение
- Продвинутая оптимизация
- Регуляризация + **Weight Decay**
- Max-norm-регуляризация
- Dropout
- Увеличение выборки + Расширение выборки (**Data Augmentation**)
- Обрезка градиентов (**Gradient clipping**)
- Доучивание уже настроенных нейросетей (**Pre-training**)
- **Unsupervised Learning**
- Техника зануления весов (разреживания НС)
- Использование специальных архитектур под задачу
(например, использующих локальность – свёрточных НС)
- зашумление (**inject noise**)

Продвинутая оптимизация

momentum – стохастический градиент с инерцией

$$\begin{aligned} m^{(t+1)} &= \rho m^{(t)} + \nabla L^{(t)}(w^{(t)}) \\ w^{(t+1)} &= w^{(t)} - \eta m^{(t+1)} \end{aligned}$$

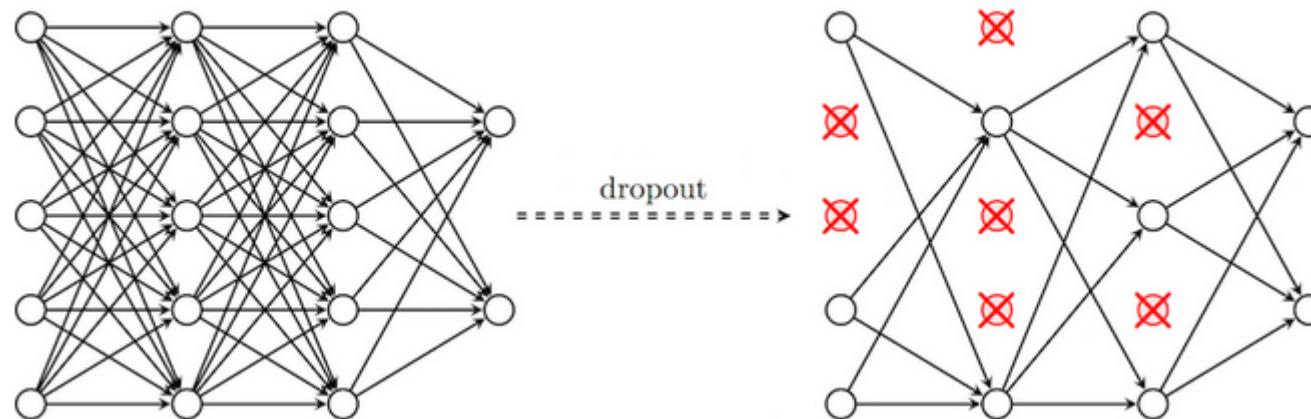
метод Нестерова

$$\begin{aligned} m^{(t+1)} &= \rho m^{(t)} + \nabla L^{(t)}(w^{(t)} - \eta m^{(t)}) \\ w^{(t+1)} &= w^{(t)} - \eta m^{(t+1)} \end{aligned}$$

Adam = RMSprop + momentum

$$\begin{aligned} m_i^{(t+1)} &= \alpha m_i^{(t)} + (1-\alpha) \nabla_i L^{(t)}(w^{(t)}) \\ v_i^{(t+1)} &= \beta v_i^{(t)} + (1-\beta) (\nabla_i L^{(t)}(w^{(t)}))^2 \\ w_i^{(t+1)} &= w_i^{(t)} - \frac{\eta}{\sqrt{v_i^{(t+1)} + \epsilon}} m_i^{(t)} \end{aligned}$$

Dropout



**случайное обнуление активаций
~ выбрасывание нейронов из сети
не выбрасываем из последнего слоя**

Обычно в полносвязных слоях!

**Отключают в режиме теста – выход умножается на вероятность
выбрасывания.**

Батч-нормализация (Batch normalization)

минибатч $\{x_i\}_{i=1}^m$

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad \text{среднее по мини-батчу}$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad \text{дисперсия по мини-батчу}$$

$$x_i^{\text{new}} = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad \text{нормировка}$$

$$y_i = \gamma x_i^{\text{new}} + \beta \quad \text{растяжение и сдвиг}$$

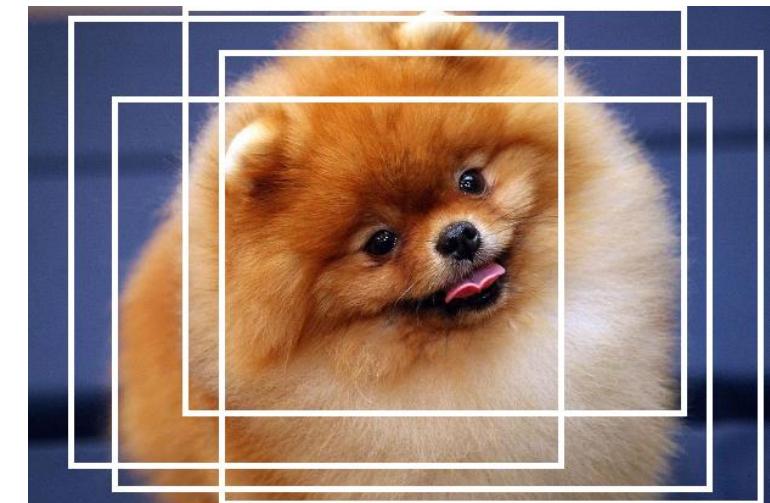
Надо определить параметры γ и β

Расширение обучающего множества (Data Augmentation)

Аугментация – построение дополнительных данных их исходных

Изображения

- **симметрии (flip)**
- **вырезки (crop)**
- **изменение масштаба (rescaling)**
- **случайные модификации (+шум)**
 - **повороты (rotation)**
 - **сдвиги (shift)**
- **изменение яркости, контраста, палитры**
 - **эффекты линзы**
- **перерисовка изображения (ex GAN)**



Звук

- **+фоновый шум**
- **тональность**

Текст

- **замена синонимов**

Тонкость: преобразования могут переводить объект в другой класс, например повороты «6» и «9».

Классификация изображений: определение породы собаки



Shiba Inu

Lively, independent-minded, aloof
toward strangers



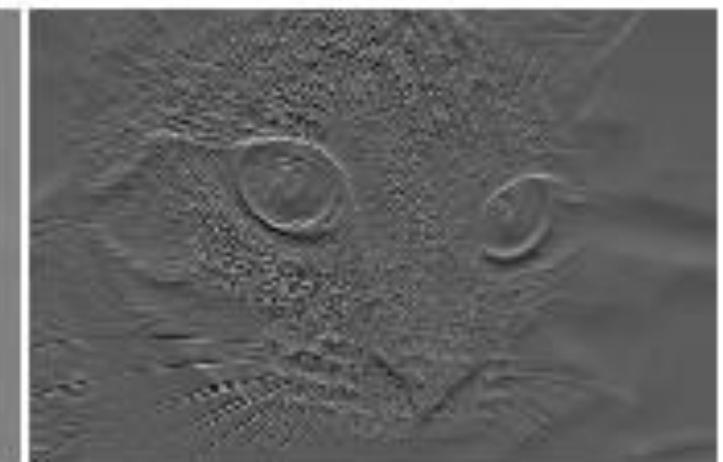
German Shepherd Dog

Must be kept busy, intense, highly
intelligent

<https://www.what-dog.net/>

**Аналогично: определение марки
авто по фото**

Классификация изображений: свёрточные НС



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

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

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

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

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

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

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

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

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

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

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

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

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

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

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

12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

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

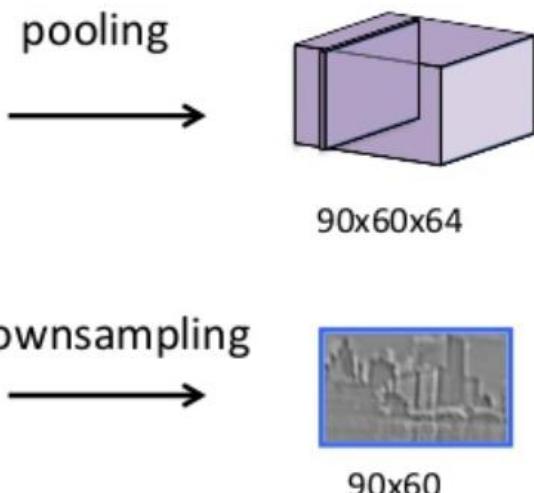
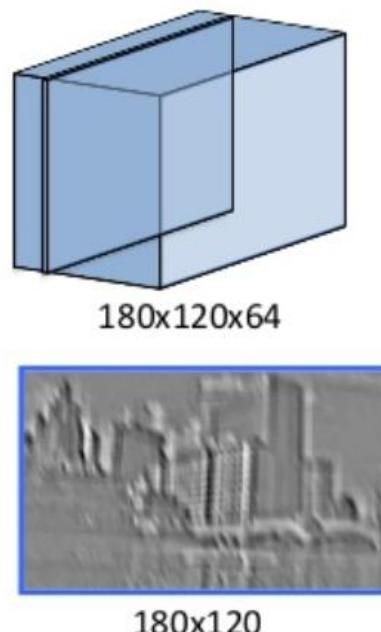
12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0

Важно**Раньше: обработка изображений – специально построенные свёртки****Сейчас: не будем специально строить свёртки – их параметры настраиваются сами!****Свёртку можно применять к изображениям любых размеров!****Pooling (агрегация, субдискретизация / subsampling)**

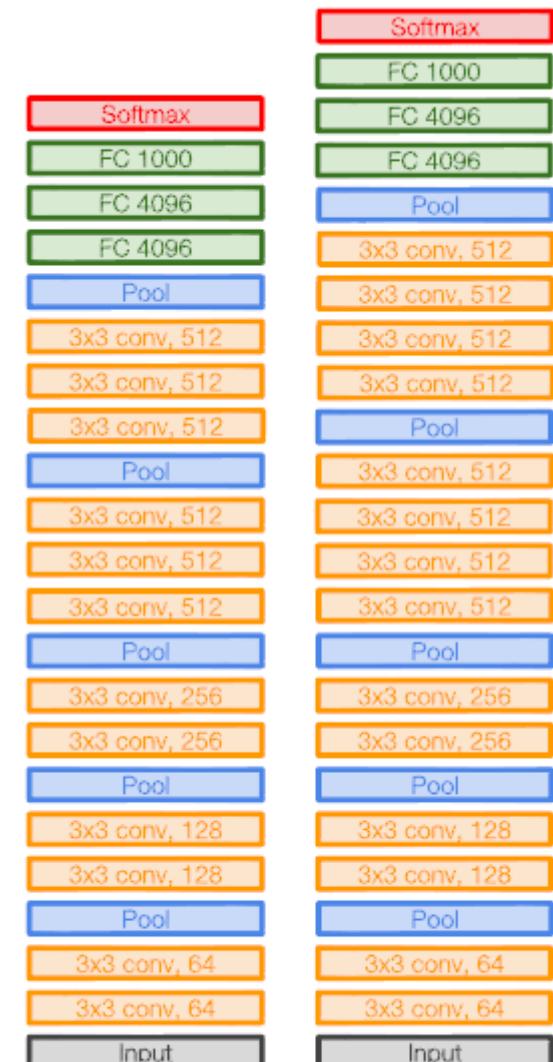
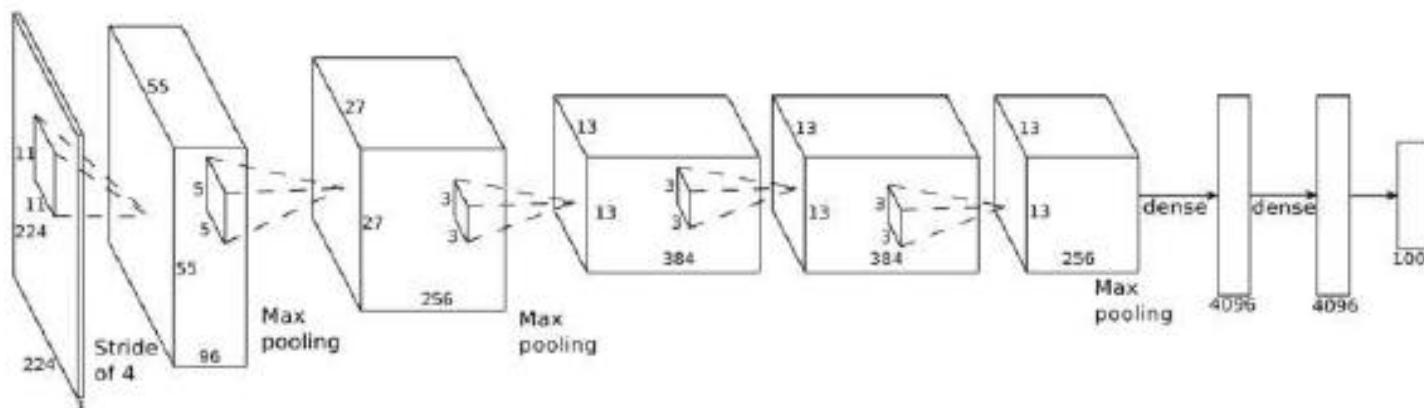
12	20	30	0
8	12	2	0
34	70	37	4
112	100	25	12

2×2 Max-Pool

20	30
112	37



VGG (2014)



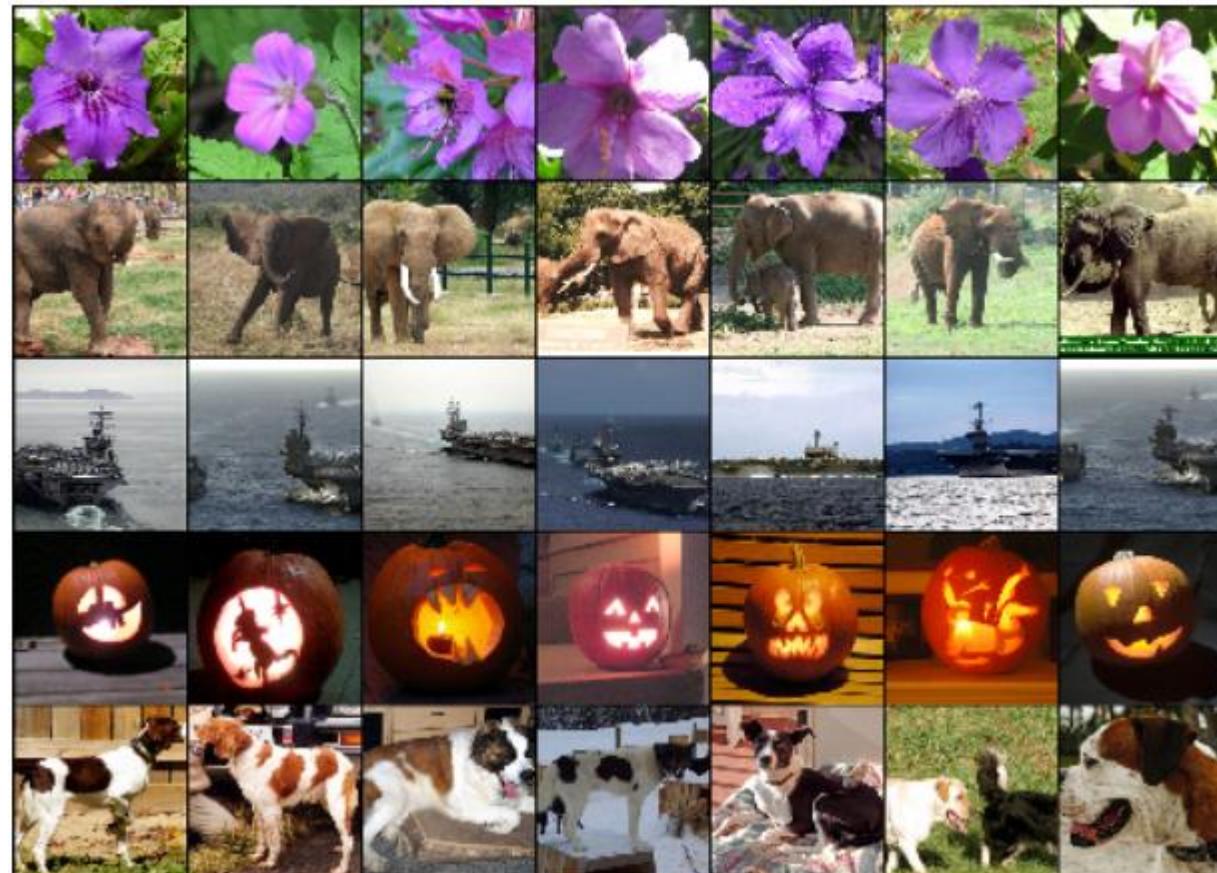
VGG16

VGG19

Визуализация – как понять, чему сеть учится...

Последний полносвязный слой

Можно смотреть соседей в этом признаковом пространстве



<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

Визуализация – как понять, чему сеть учится...

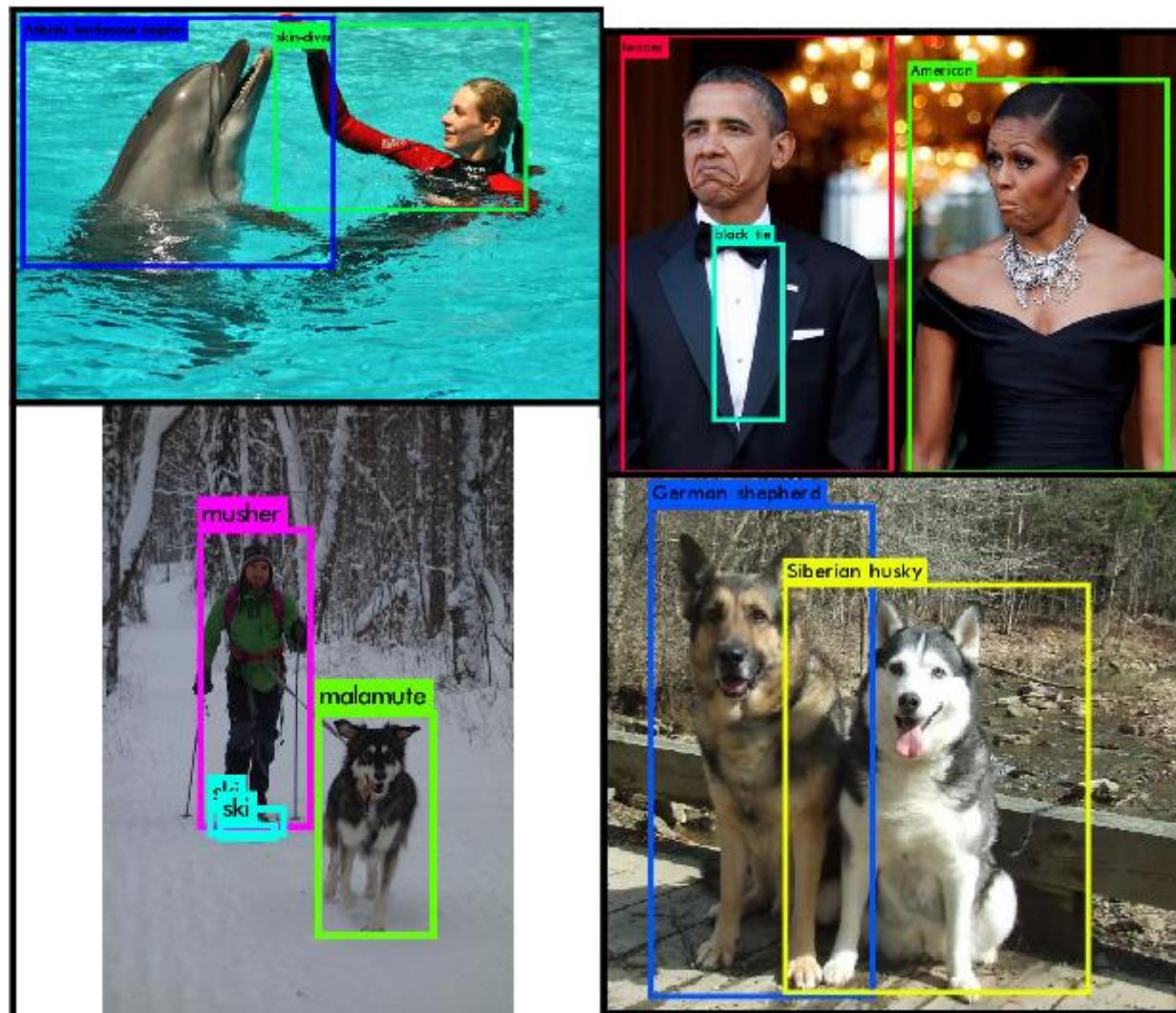
Средние слои

На каких изображениях максимальные значения активаций



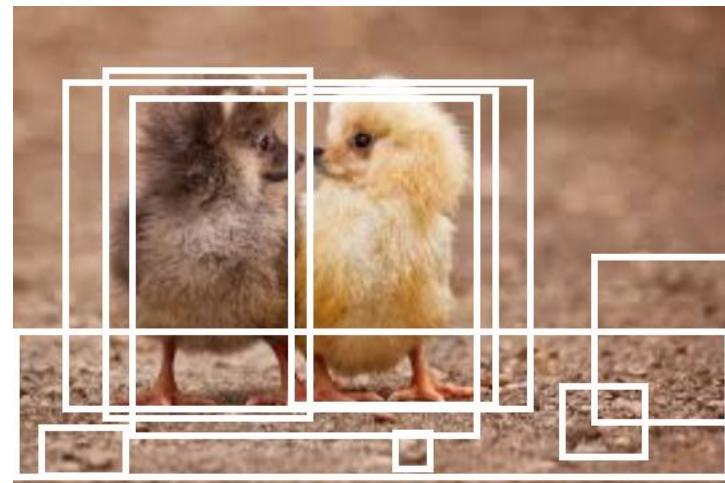
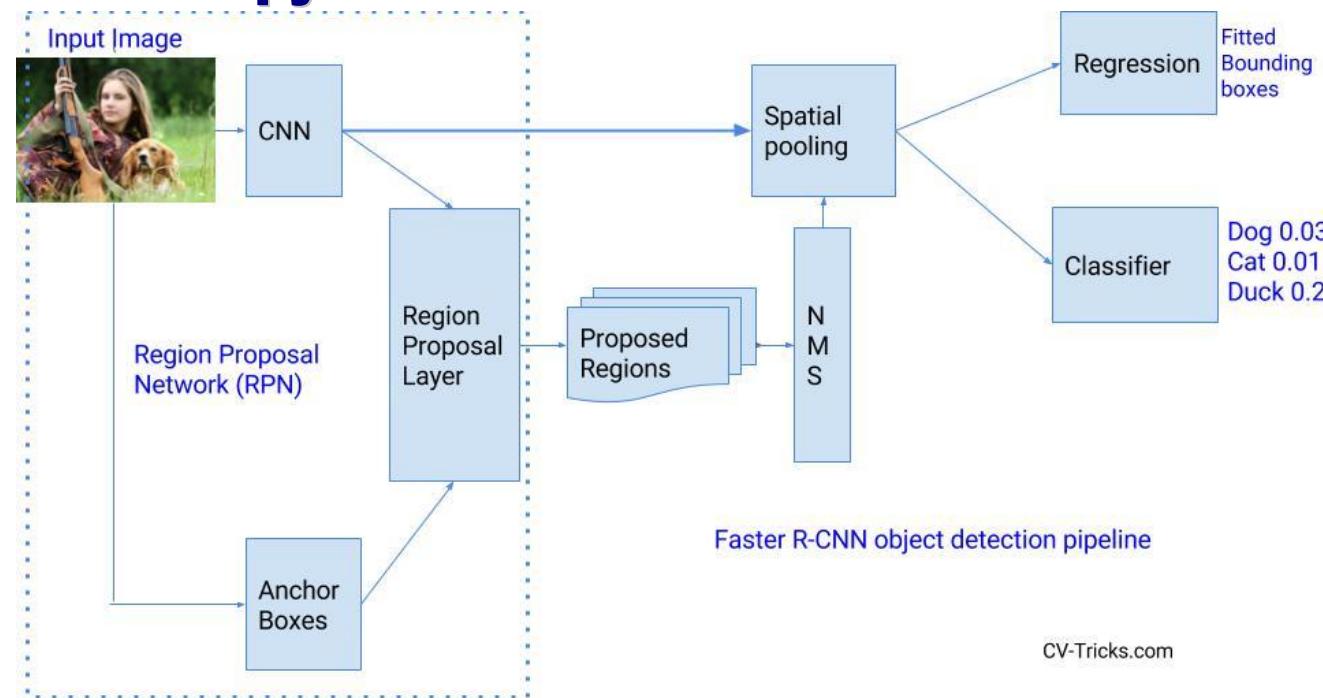
<https://arxiv.org/pdf/1412.6806.pdf>

Обнаружение объектов (Object Detection)



Joseph Redmon, Ali Farhadi «YOLO9000: Better, Faster, Stronger» <https://arxiv.org/abs/1612.08242>

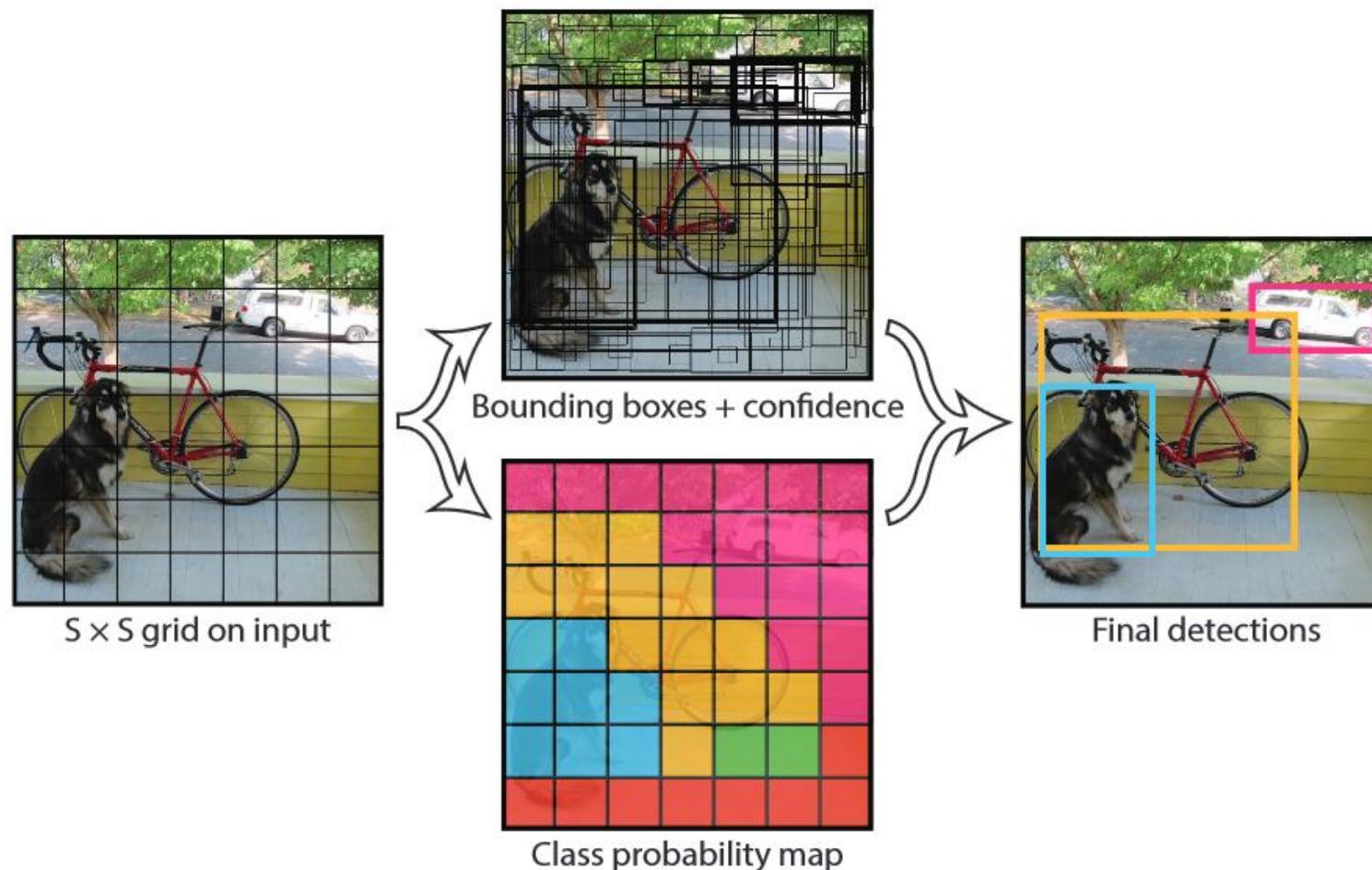
Обнаружение объектов: Faster R-CNN



Обнаружение объектов: YOLO (You only Live Look Once)

- Вход: Изображение (изменение масштаба → 448×448)

Выход: класс (вероятности) + координаты рамки



Сегментация объектов

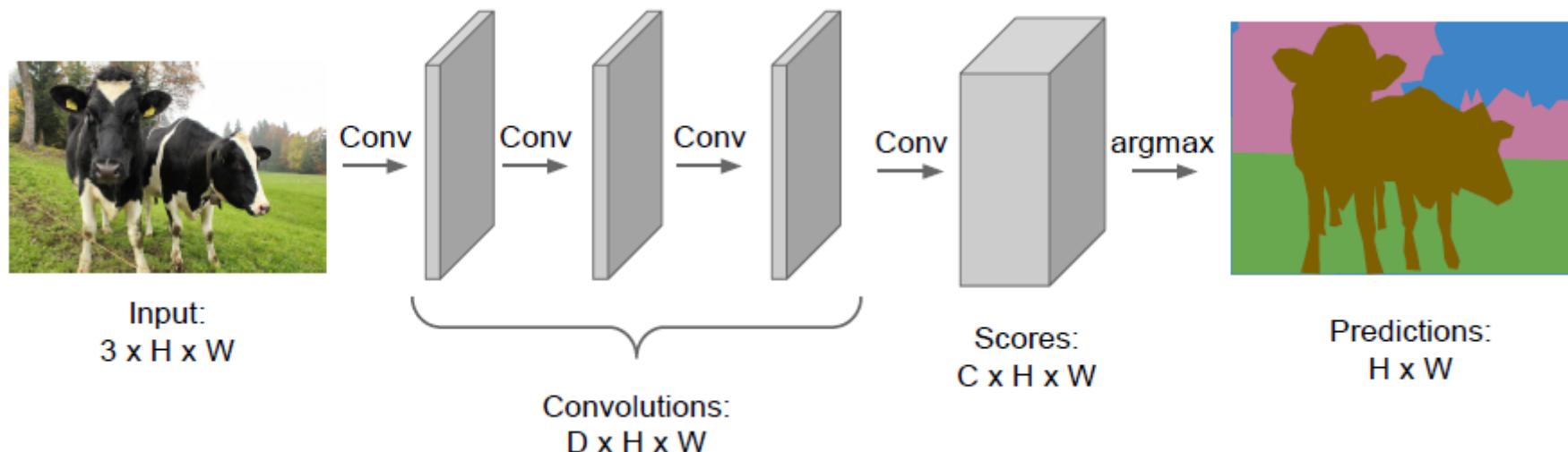


Не и др., 2012

<https://research.fb.com/learning-to-segment/>

Сегментация объектов

**Идея: полностью свёрточная структура сети –
Fully Convolutional Network (FCN)**

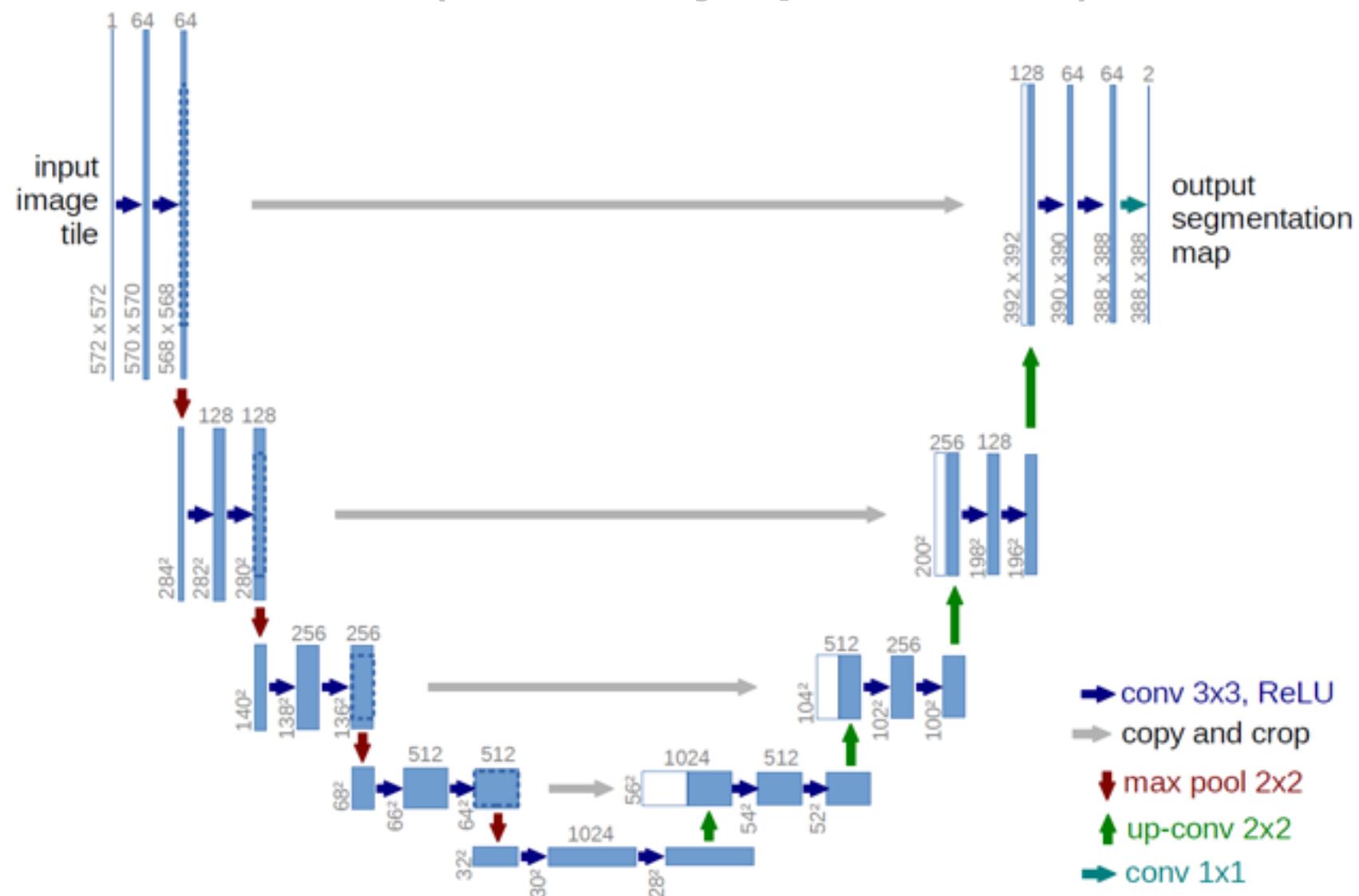


Проблемы: высокое разрешение исходного изображения

- надо его понизить (**downsampling**)
 - **pooling**
 - **strided convolution**
- потом надо восстановить (**upsampling**)

<http://cs231n.stanford.edu>

U-Net (самая популярная сейчас)



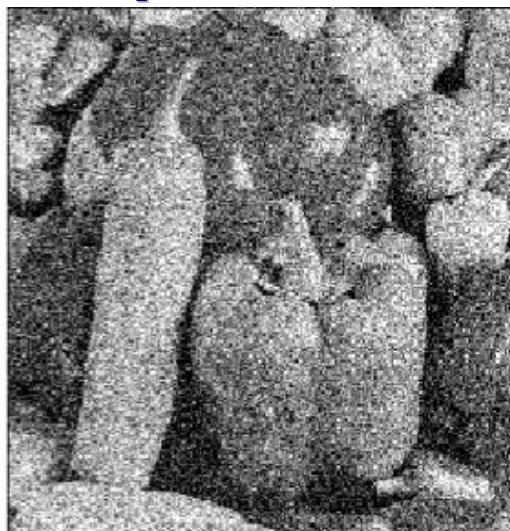
«U-Net: Convolutional Networks for Biomedical Image Segmentation» [Ronneberger O. и др., 2015
<https://arxiv.org/abs/1505.04597>】

FCN: Удаление фона (Background removal)



<https://towardsdatascience.com/background-removal-with-deep-learning-c4f2104b3157>

FCN: Улучшение изображений: Denoising / Restoring



FCN: Inpainting

nd Sirius form a nearly equilateral triangle. These stars, and Naos, in the Ship, and Phaet, in the Dove, form a figure known as the Egyptian "X." From earliest times Sirius has been known as the Dog of Orion. It is 324 times brighter than the average sixth-magnitude star, and is the nearest star to our earth of all the stars in this latitude; its distance being 8.7 light years. At this distance the Sun would appear a little brighter than the Pole Star. [Illustration CANIS MAJOR] — ARGO NAVIS (Аргонавтика).—ARGO. (Face South.) — LOCATION.—Argo is situated in Canis Major. If a line joining Betelgeuze and Sirius is prolonged 18° southeast, it will point out Naos, a second magnitude in the rowlock of the Ship. This is in the southeast corner of the Egyptian "X." The star of a deep yellow or orange hue. It has three little stars above it, two of which form a pretty pair. The star Pi is a companion, which is a test for an opera-glass. The star Sigma is a double for an opera-glass. Note the fine star cluster Mu. The star Markeb forms a small triangle with two stars near it. The Egyptians believed that this was the star that bore Osiris and Isis over the Deluge. The constellations contains two noted objects invisible in our latitude, Canopus, the second brightest star, and the remarkable variable star Iota. [Illustration PUPPIS] — MONOCEROS (монохор).—MONOCEROS (Face South.) — LOCATION.—Monoceros is to be found east of Orion between Canis Minor. Three of its stars of the fourth magnitude lie in a straight line northeast and southwest, about 9.4° apart. Betelgeuze, and about the same distance south of Aldebaran, the region around the stars 8, 13, 17 is particularly rich when viewed with an opera-glass. Note also a field about the variable Sigma, and a cluster about midway between the stars 12 and 13. Two stars about 7° apart in the tail of the Unicorn are winter stars to Procyon. These stars are

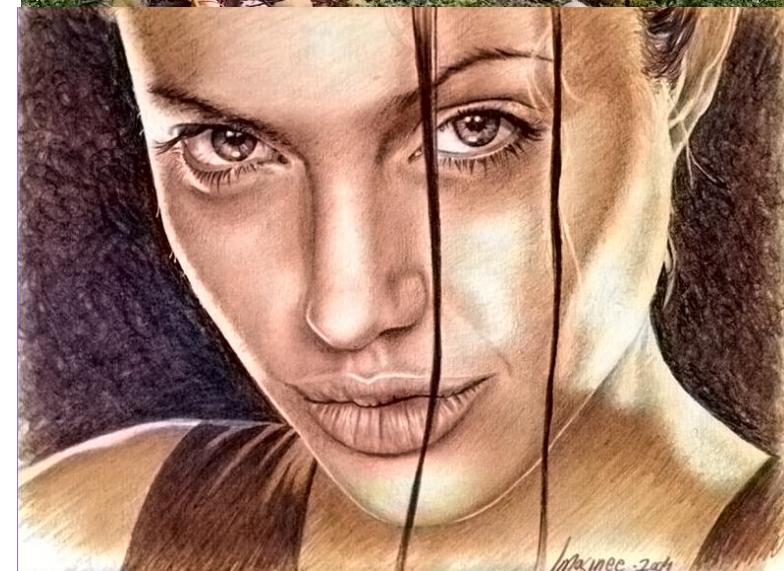


FCN: Дорисовка изображений (Image completion)



http://hi.cs.waseda.ac.jp/~iizuka/projects/completion/data/completion_sig2017.pdf

FCN: Раскрашивание изображений



<http://demos.algorithmia.com/colorize-photos/>

в 20% случаях раскрашенное изображение путали с оригиналом...

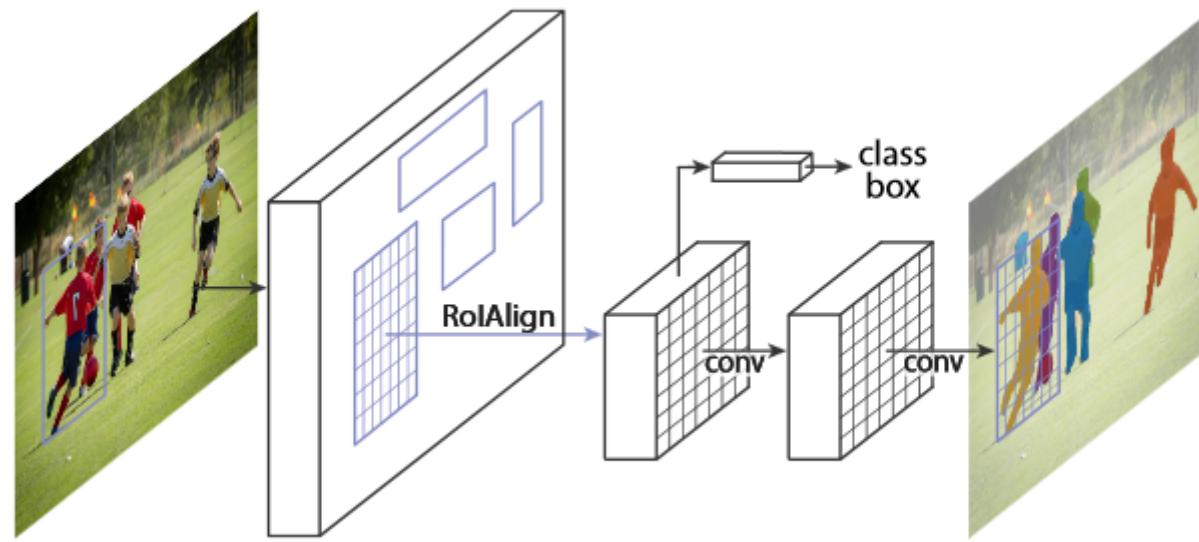
Сегментация объектов

не просто найти рамку, а выделить объект!



Mask R-CNN <https://arxiv.org/pdf/1703.06870.pdf>

Сегментация объектов: Mask R-CNN

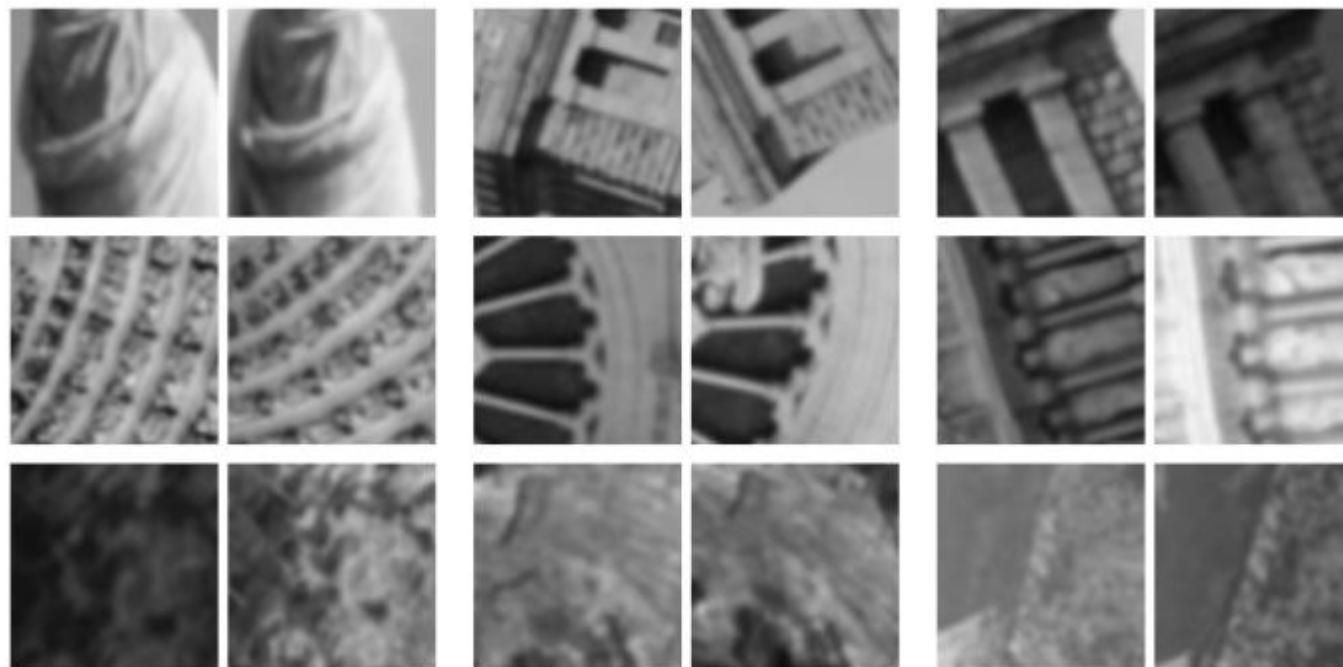
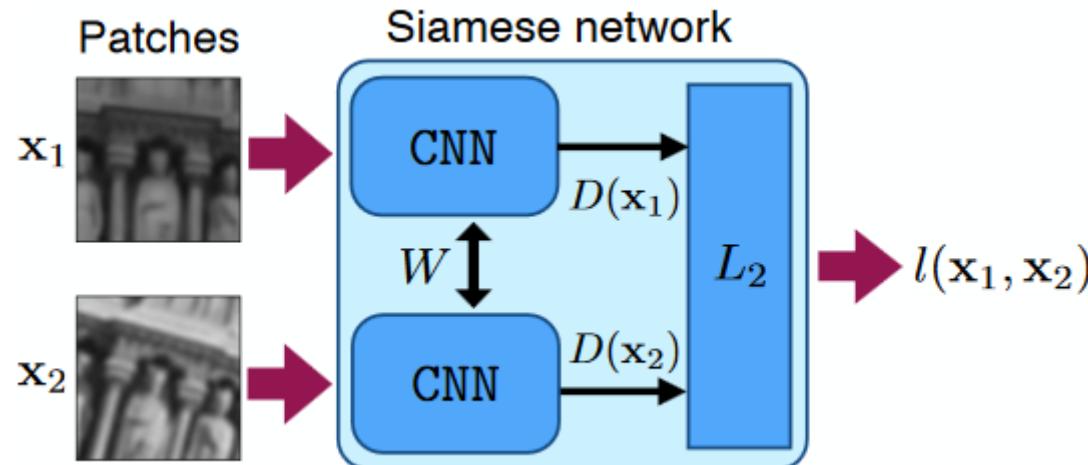


**Faster R-CNN + определение сегментационной маски
(маленькая FCN)**

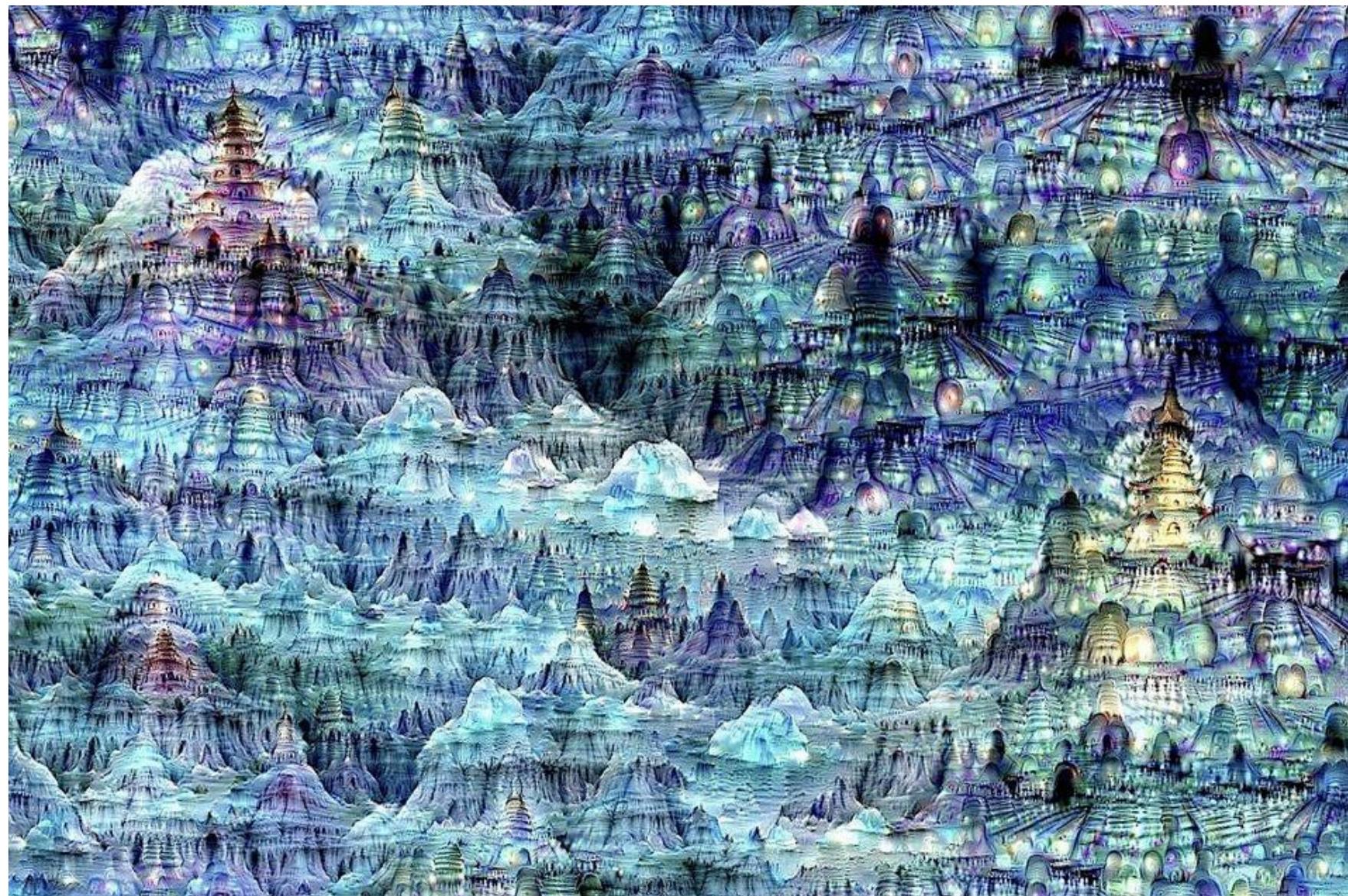
+ борьба за более точное определение границ

Поиск изображений: сиамские НС

Идея: изображение → вектор, поиск по векторам



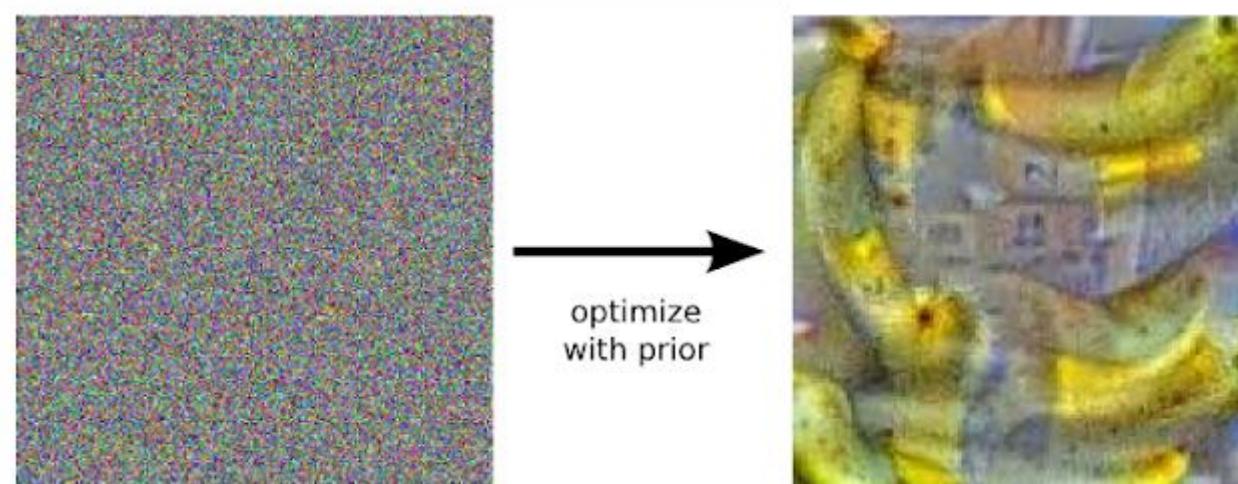
Генерация изображений



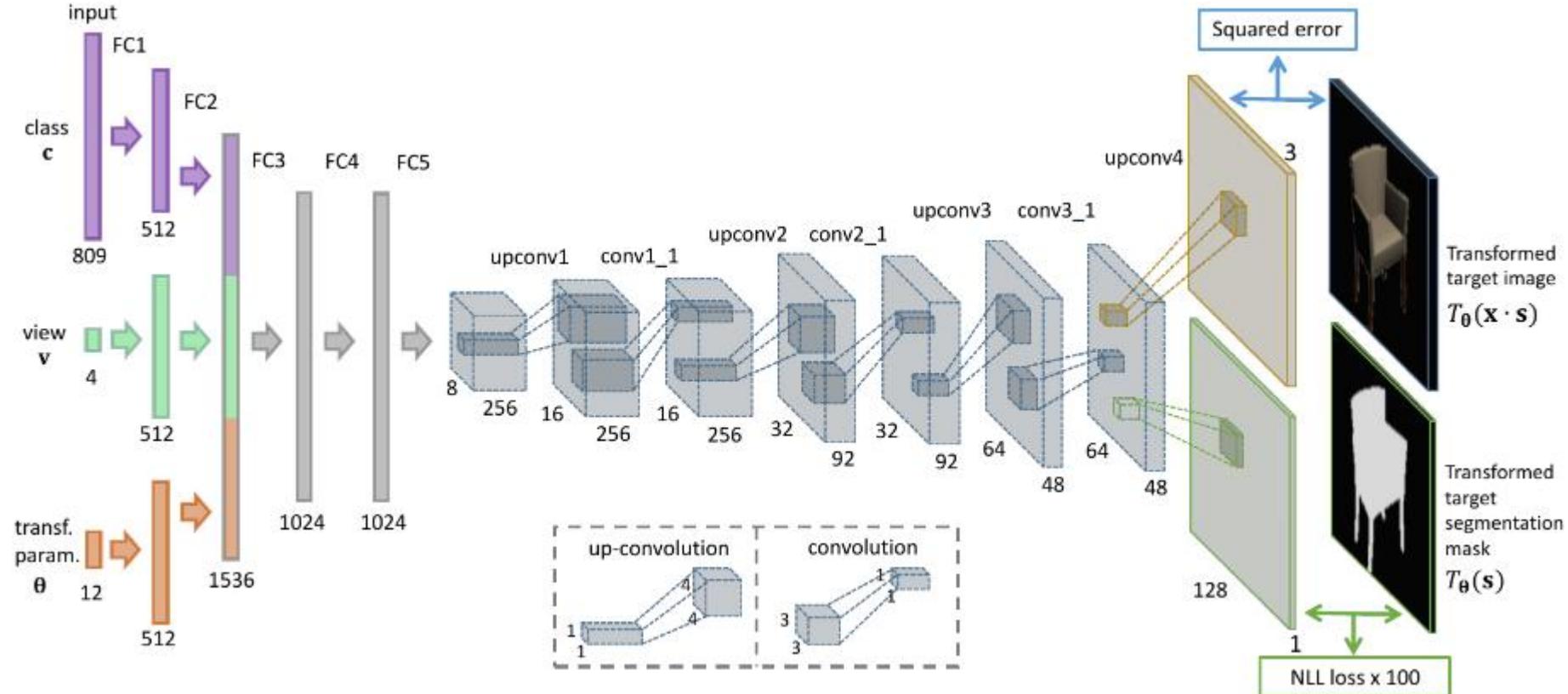
Генерация изображений

**Натренировали НС-классификатор
Дали на вход случайный шум и стали градиентно менять
изображение: увеличить уверенность, что это «банан»**

Есть хак: соседние пиксели коррелированы



Генерация изображений по стилю, ориентации (относительно камеры), цвету, яркости и т.п.



как бы развёрнутая CNN

параметры → 1024 признака → изображение и маска сегментации

Dosovitskiy A. и др. «Learning to Generate Chairs, Tables and Cars with Convolutional Networks», 2017 // <https://arxiv.org/pdf/1411.5928.pdf>

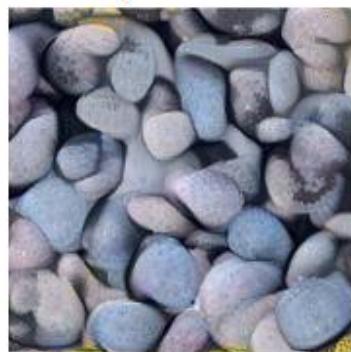
Генерация объектов / представление объектов



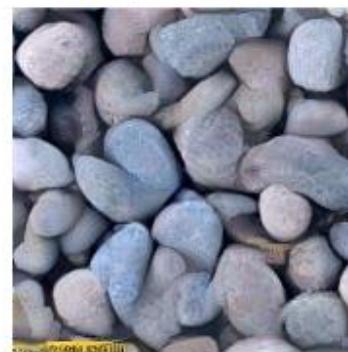
[Dosovitskiy A. и др., 2017 <https://arxiv.org/pdf/1411.5928.pdf>]

Генерация текстур

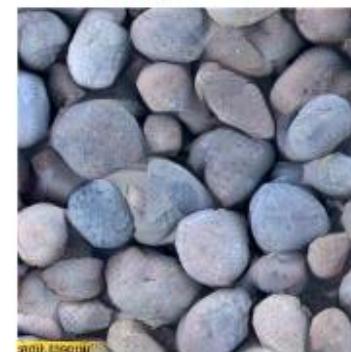
A ~1k parameters



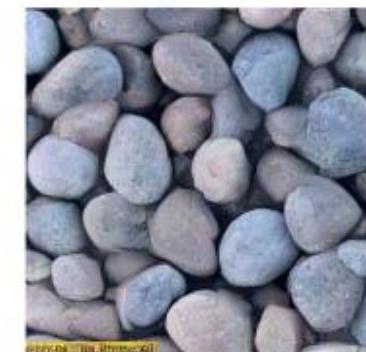
~10k parameters



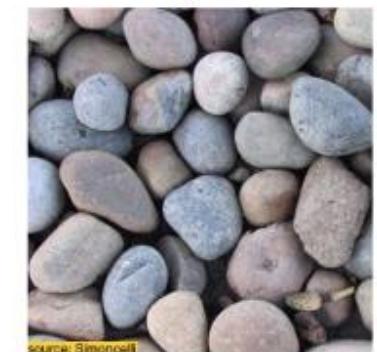
~177k parameters



~852k parameters

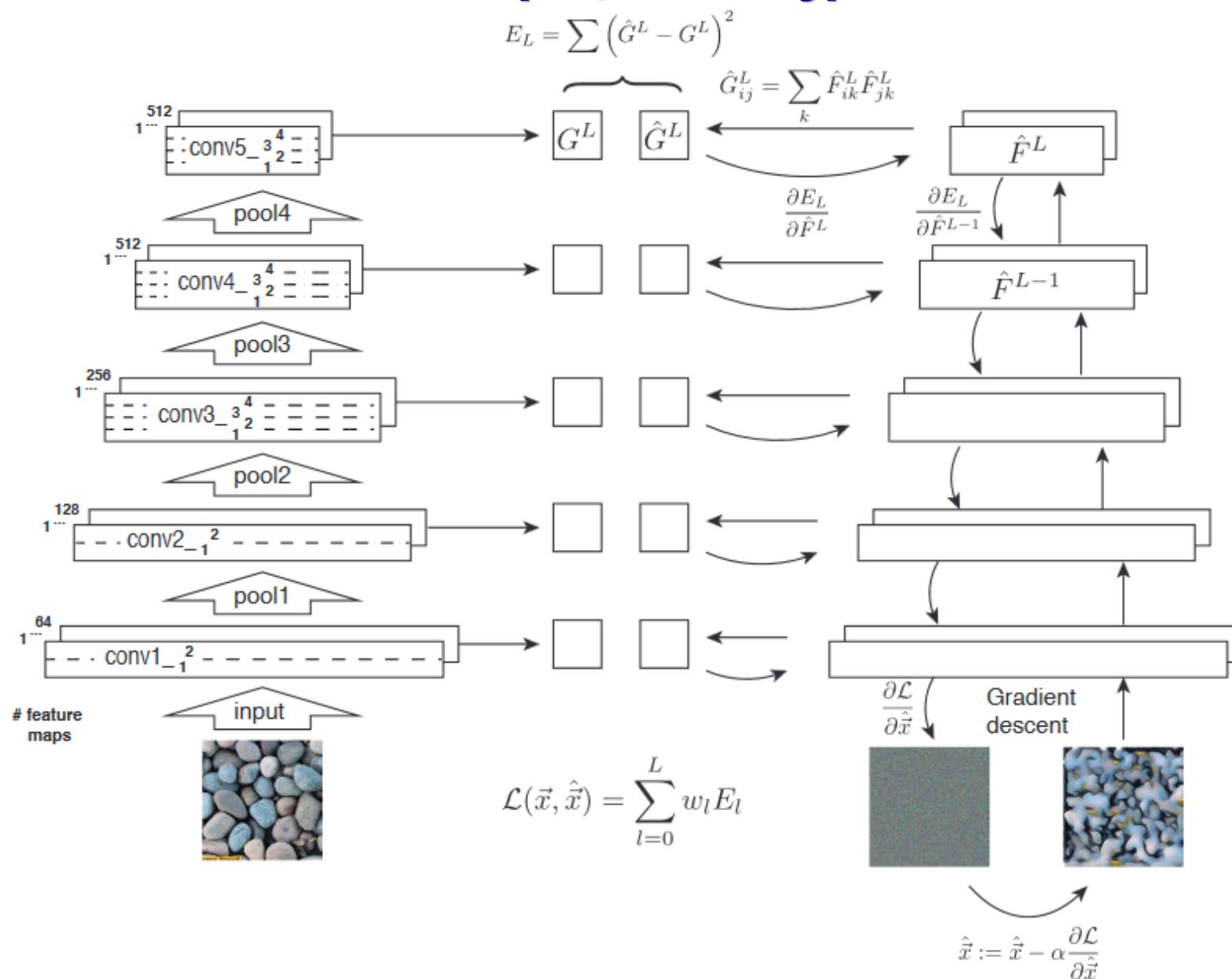


original

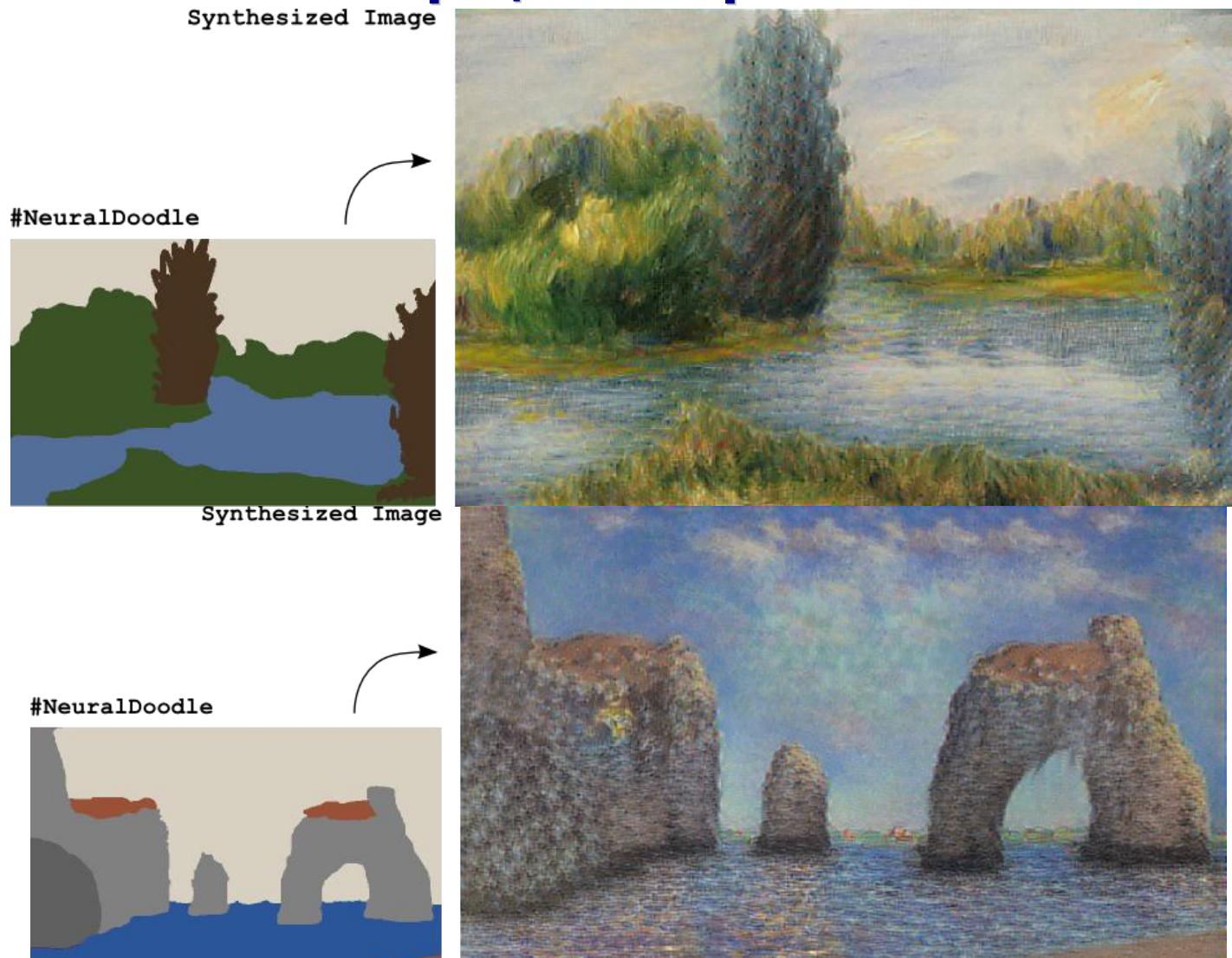


Leon A. Gatys «Texture Synthesis Using Convolutional Neural Networks» <https://arxiv.org/pdf/1505.07376.pdf>

Генерация текстур



Генерация изображений



<https://github.com/alexjc/neural-doodle>

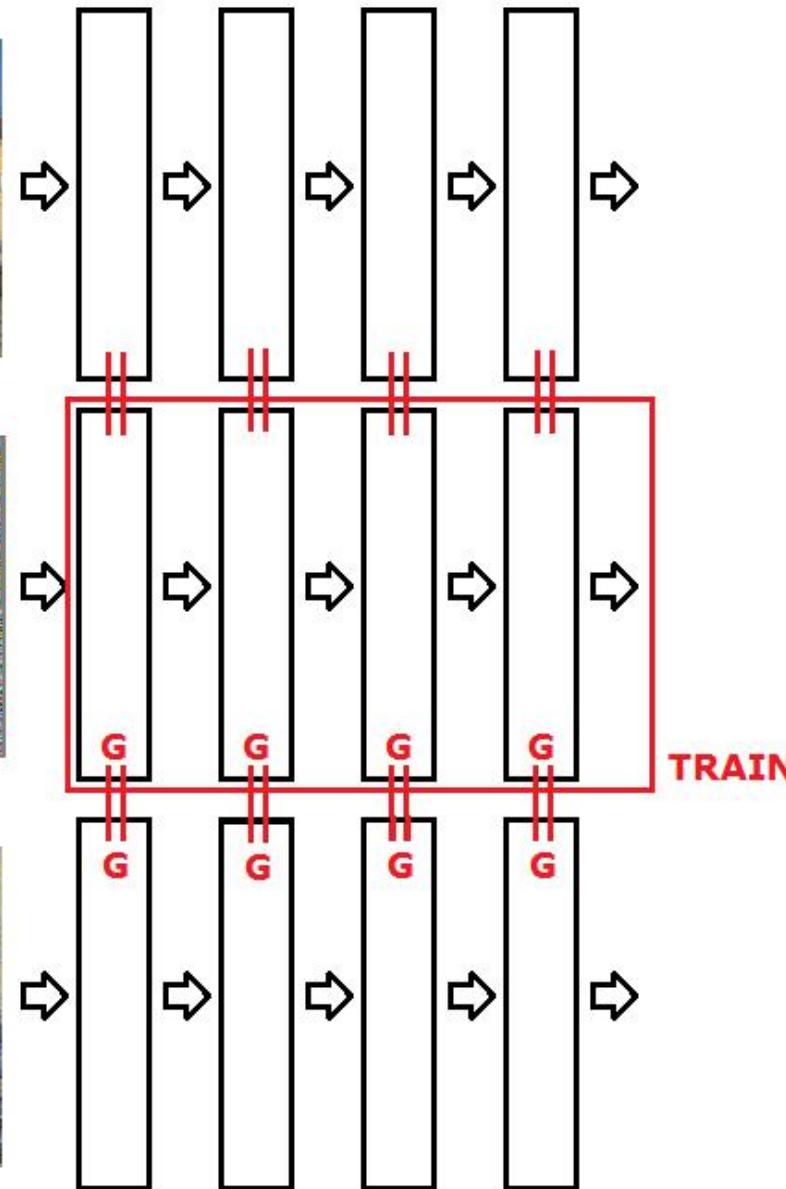
Стилизация изображений, перенос стиля



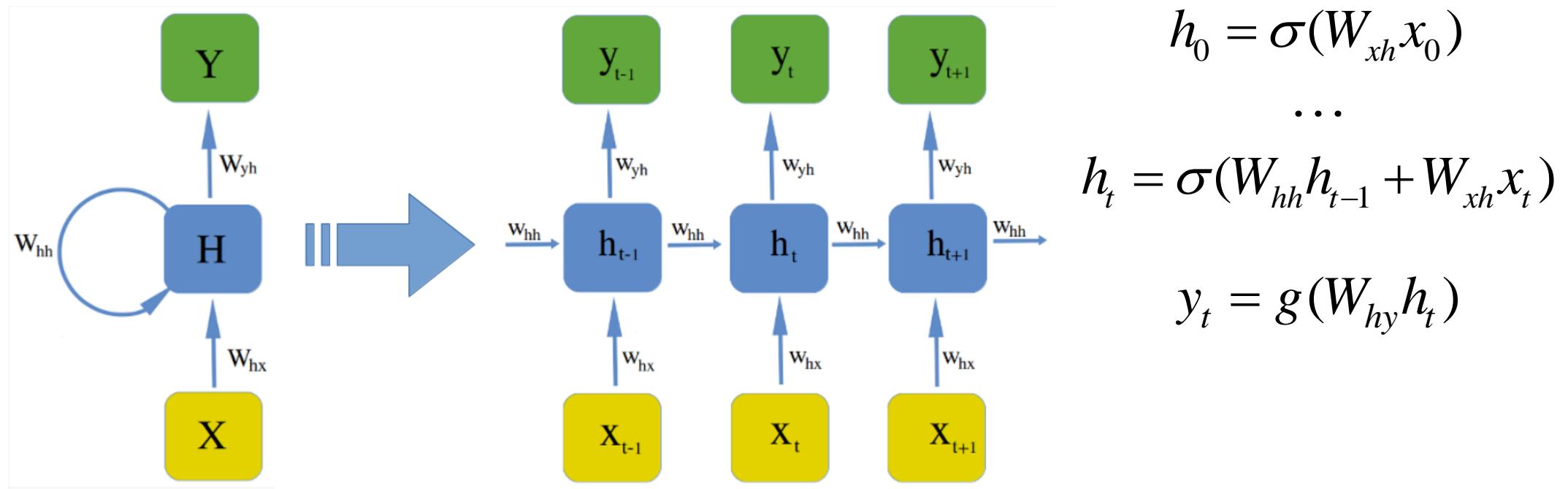
<https://deepart.io/>

<https://arxiv.org/pdf/1703.07511.pdf>

Стилизация



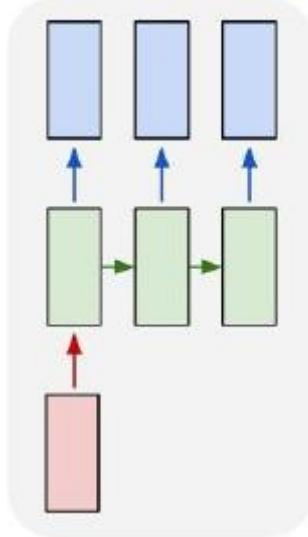
Генерация текста: (RNN = Recurrent neural network)



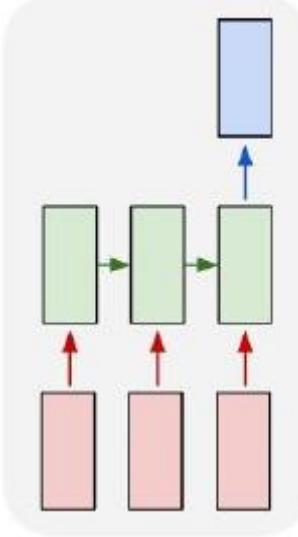
<http://www.jefkine.com/general/2018/05/21/2018-05-21-vanishing-and-exploding-gradient-problems/>

Применение RNN

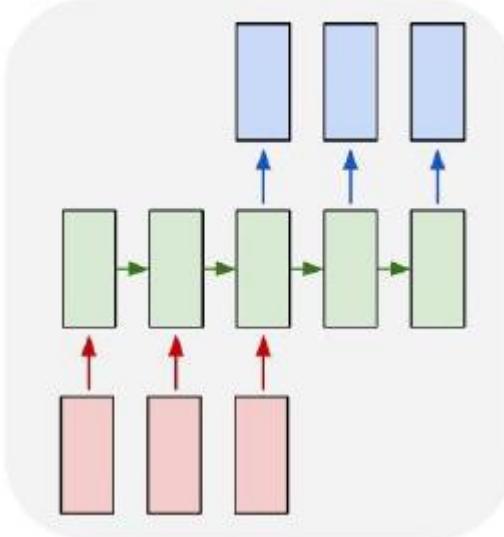
one to many



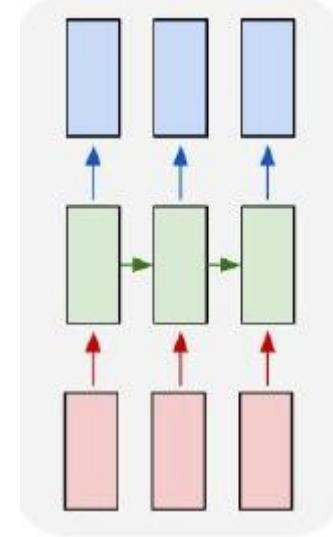
many to one



many to many



many to many



**описание
изображения**

**тема /
настроение
текста**

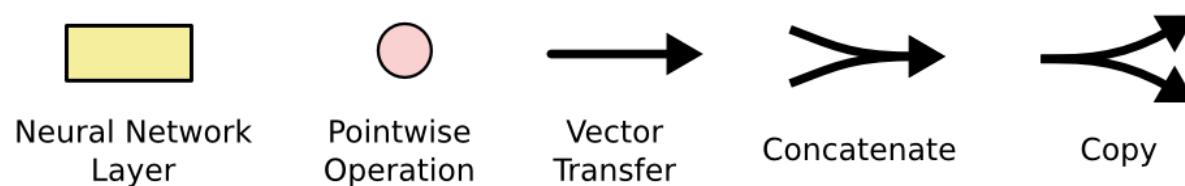
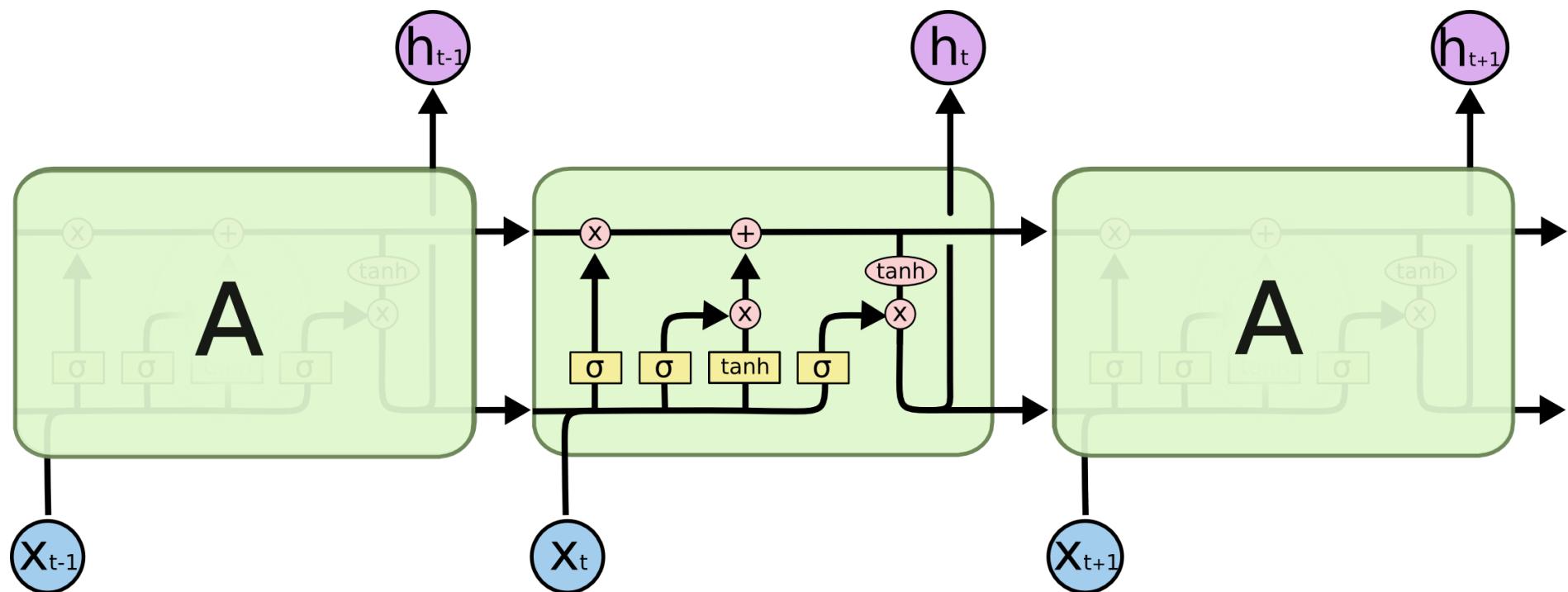
**машинный
перевод**

**классификация
фреймов видео**

**Можно по-разному собирать блоки –
для решения разных задач**

<http://karpathy.github.io/2015/05/21/rnn-effectiveness/>

LSTM (другой базовый блок)



[Hochreiter&Schmidhuber, 1997]

Интерпретация RNN

Cell sensitive to position in line:

The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action--the one Kutuzov and the general mass of the army demanded--namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all--carried on by vis inertiae--pressed forward into boats and into the ice-covered water and did not, surrender.

Cell that turns on inside quotes:

"You mean to imply that I have nothing to eat out of... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

**Отдельные нейроны – «счётчики числа слов в предложении»,
«индикатор – текст в кавычках»**

**Karpathy, Johnson, and Fei-Fei: Visualizing and Understanding Recurrent Networks, ICLR
Workshop 2016 <http://vision.stanford.edu/pdf/KarpathyICLR2016.pdf>**

Аннотирование изображений

I think it's a dog sitting on a rock.



<https://arxiv.org/pdf/1411.4555.pdf>

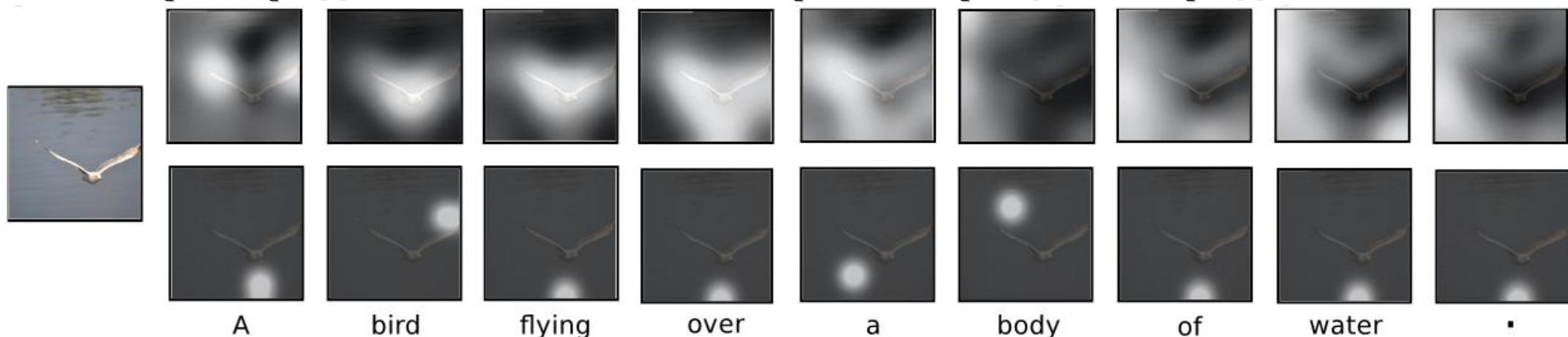
I am not really confident, but I think it's a young boy holding a baseball bat and he seems 😊.



<https://www.captionbot.ai/>

Image Captioning with Attention

Как распределяется внимание при генерации очередного слова

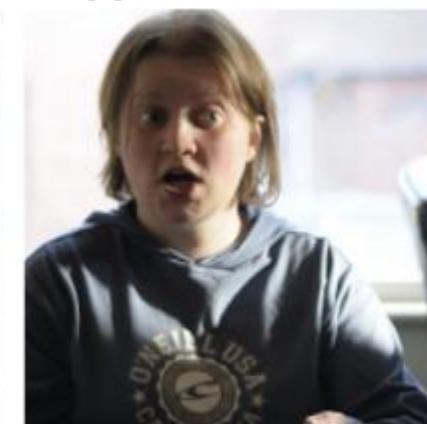


Верхний ряд – большая область внимания, нижний – маленькая

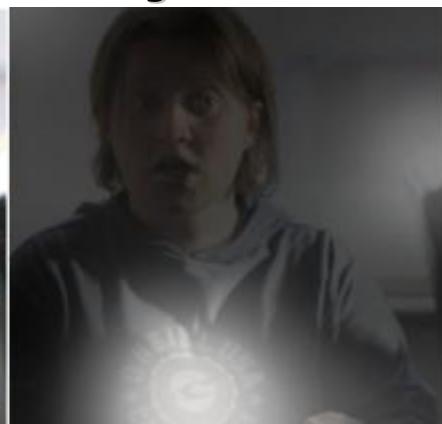
Заодно инструмент «что модель видит»... – и почему ошибается...



A large white bird standing in a forest.



A woman holding a clock in her hand.



Текст в изображение (text2image)

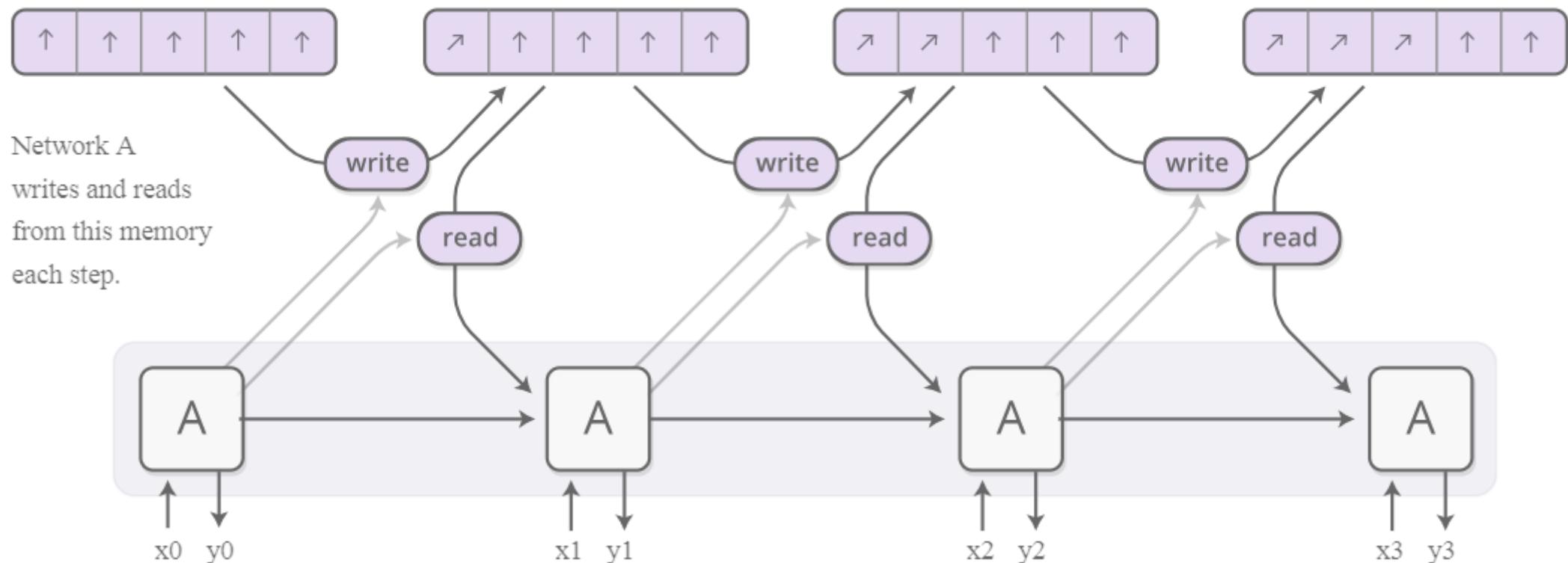
Text descriptions (content)	Images (style)
The bird has a yellow breast with grey features and a small beak.	
This is a large white bird with black wings and a red head .	
A small bird with a black head and wings and features grey wings.	
This bird has a white breast , brown and white coloring on its head and wings, and a thin pointy beak.	
A small bird with white base and black stripes throughout its belly, head, and feathers.	
A small sized bird that has a cream belly and a short pointed bill.	
This bird is completely red .	
This bird is completely white .	
This is a yellow bird. The wings are bright blue .	

S.Reed «Generative Adversarial Text to Image Synthesis», 2016

<https://arxiv.org/pdf/1605.05396.pdf>

Neural Turing Machines

Memory is an array of vectors.



<https://distill.pub/2016/augmented-rnns/#neural-turing-machines>

A. Graves, G. Wayne, I. Danihelka «Neural Turing Machines», 2014 // <https://arxiv.org/abs/1410.5401>

Анализ текстов: word2vec

Трюк: настраиваем модель, но не для использования в задаче, которой учим (нас интересуют формируемые внутренние представления) Аналогично было в автокодировщиках;)

Термины «distributional semantics»

Смысл слова определяется контекстом

Полосатая маленькая *** мурлычит и пьёт молоко**

Весна

Ручьи

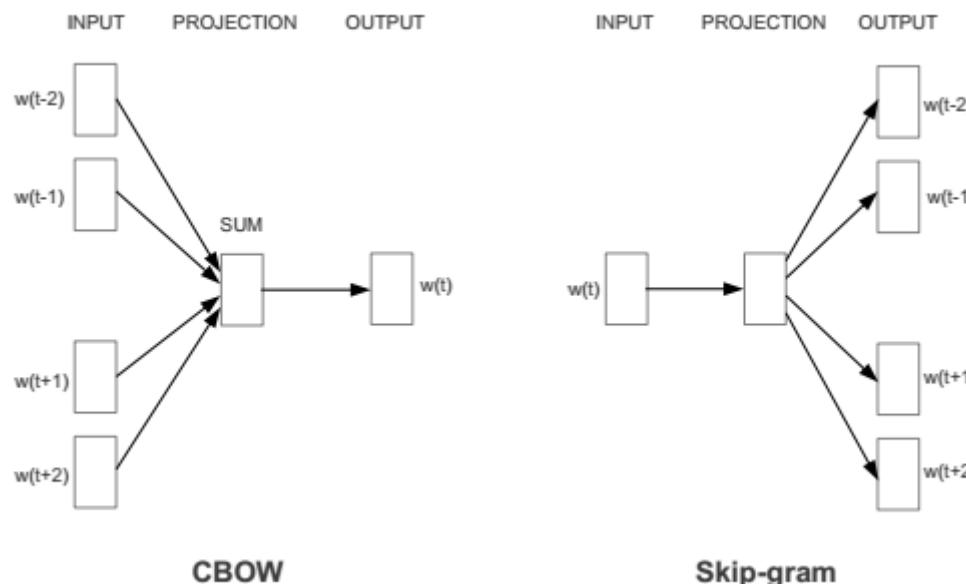
Тает

Цветёт

Зеленеет

Прилетают

word2vec: представления контекста



CBOW = Continuous Bag of Words (быстрее, окно ~ 5, большие корпуса)
skipgram model (лучше, окно ~ 10, небольшие корпуса)

Peace
Peaceful
Friendship
Nonviolence

Path
Paths
Approach
Titled
Pathway
Way

Stop
Quit
Stopped
Avoid
Resist

http://bionlp-www.utu.fi/wv_demo/

Операции над представлениями слов

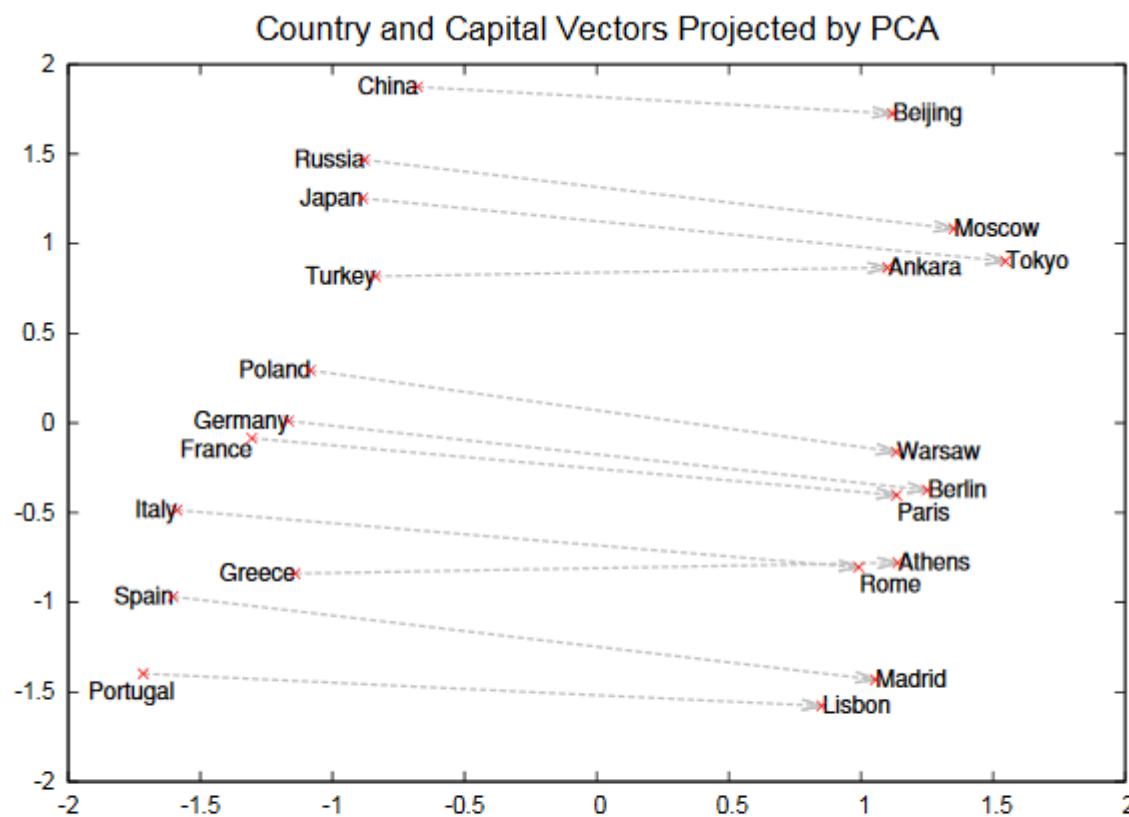
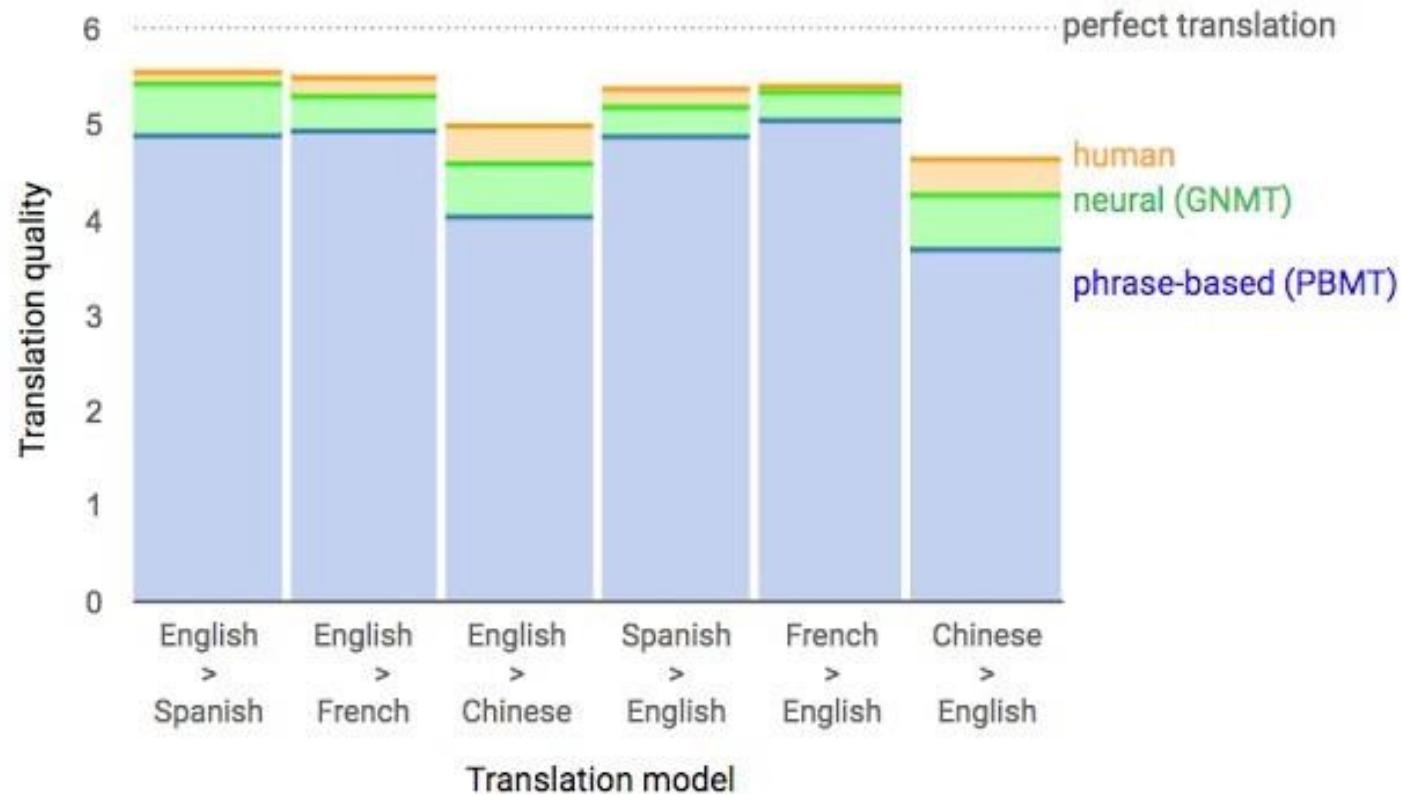


Figure 2: Two-dimensional PCA projection of the 1000-dimensional Skip-gram vectors of countries and their capital cities. The figure illustrates ability of the model to automatically organize concepts and learn implicitly the relationships between them, as during the training we did not provide any supervised information about what a capital city means.

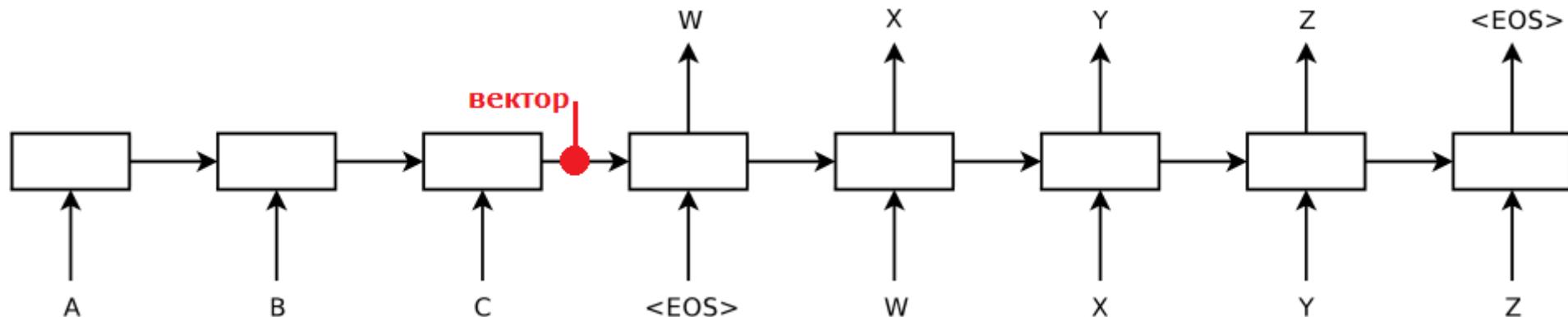
[Mikolov et al., 2013] <https://arxiv.org/pdf/1310.4546.pdf>

Автоматический перевод

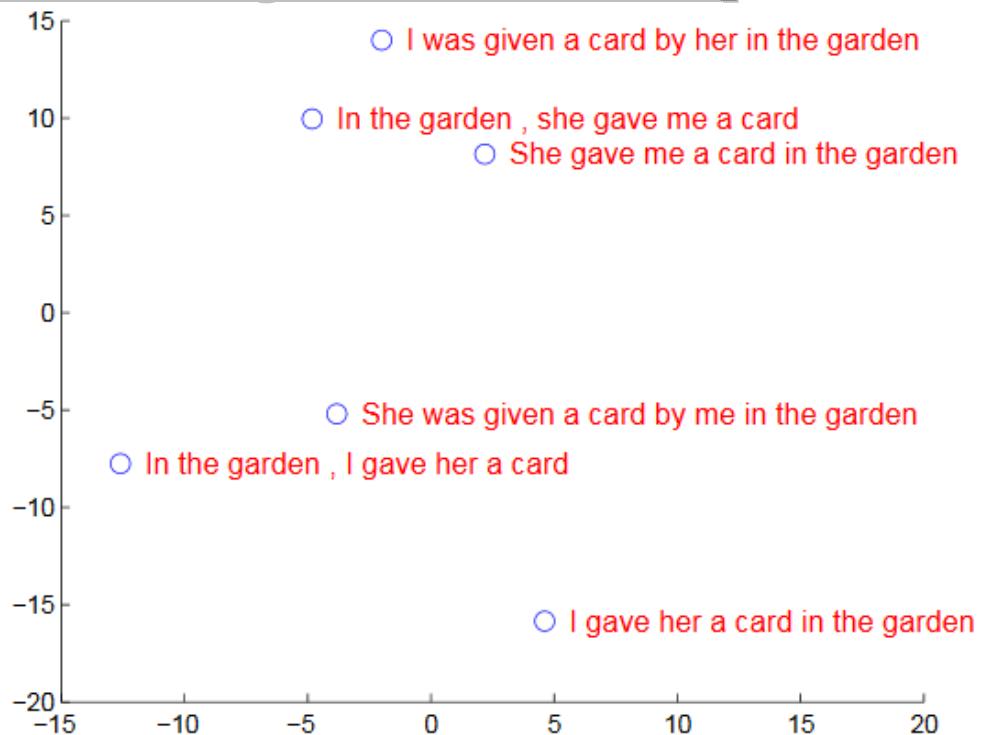
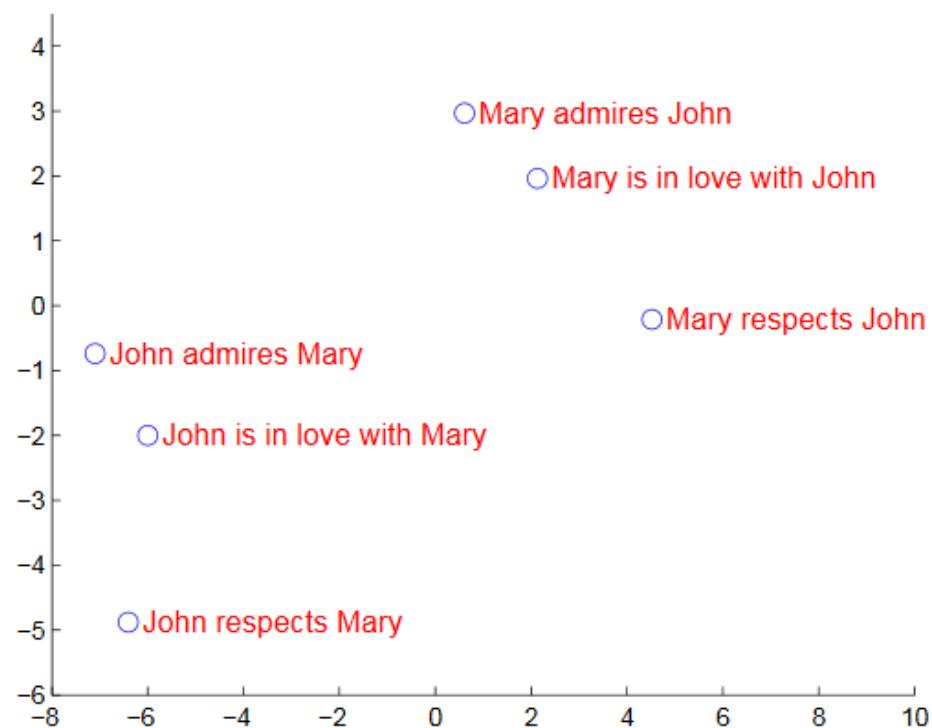


Google's Neural Machine Translation System
<https://arxiv.org/abs/1609.08144>

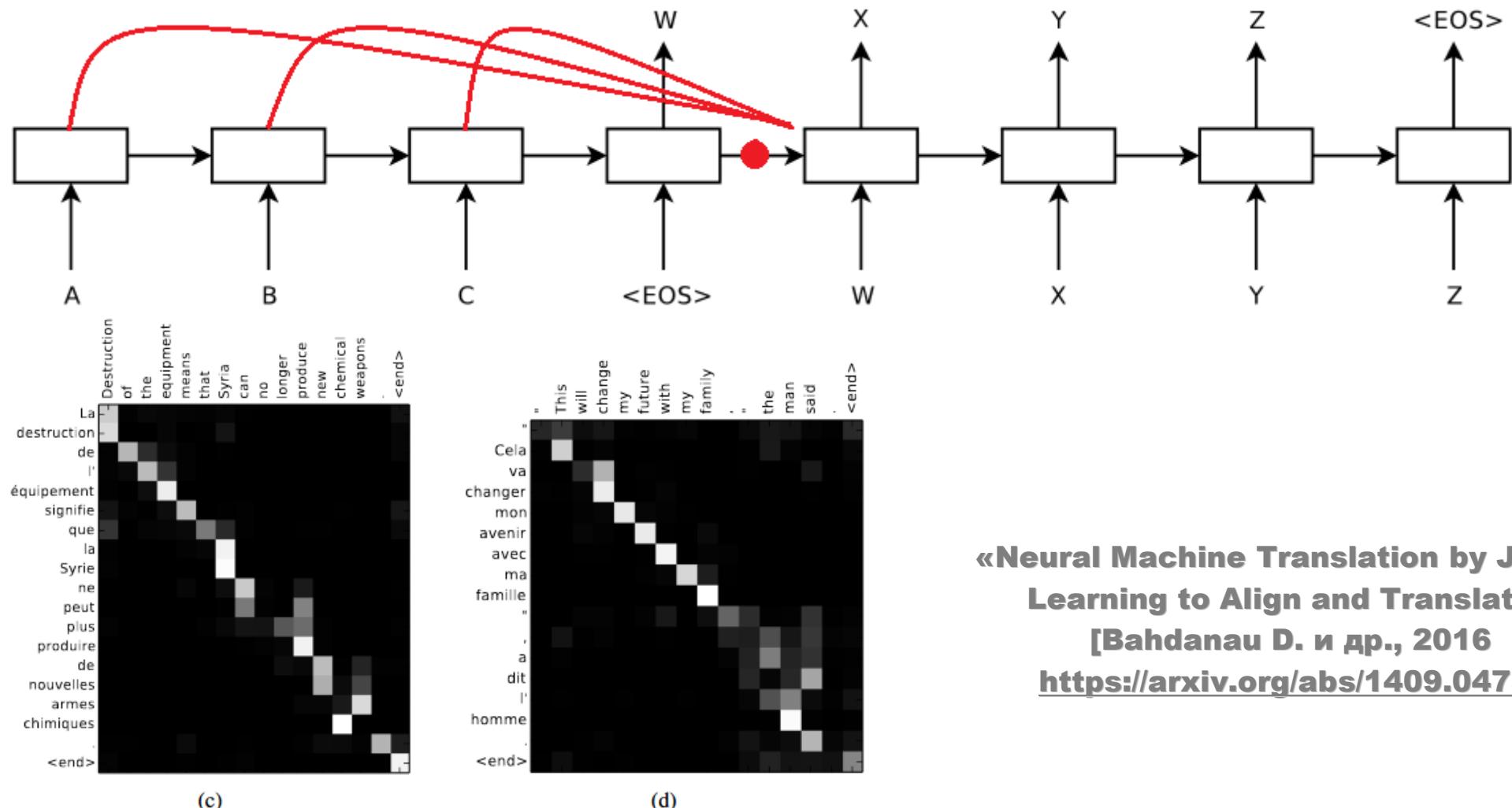
Автоматический перевод: seq2seq



«Sequence to Sequence Learning with Neural Networks»
[Sutskever I. и др. 2014, <https://arxiv.org/abs/1409.3215>]



Автоматический перевод: механизм внимания



«Neural Machine Translation by Jointly Learning to Align and Translate»
 [Bahdanau D. и др., 2016
<https://arxiv.org/abs/1409.0473>]

Figure 3: Four sample alignments found by RNNsearch-50. The x-axis and y-axis of each plot correspond to the words in the source sentence (English) and the generated translation (French), respectively. Each pixel shows the weight α_{ij} of the annotation of the j -th source word for the i -th target word (see Eq. (6)), in grayscale (0: black, 1: white). (a) an arbitrary sentence. (b-d) three randomly selected samples among the sentences without any unknown words and of length between 10 and 20 words from the test set.

Чатботы /экспертные системы: механизмы памяти

Sam walks into the kitchen.

Sam picks up an apple.

Sam walks into the bedroom.

Sam drops the apple.

Q: Where is the apple?

A. Bedroom

Brian is a lion.

Julius is a lion.

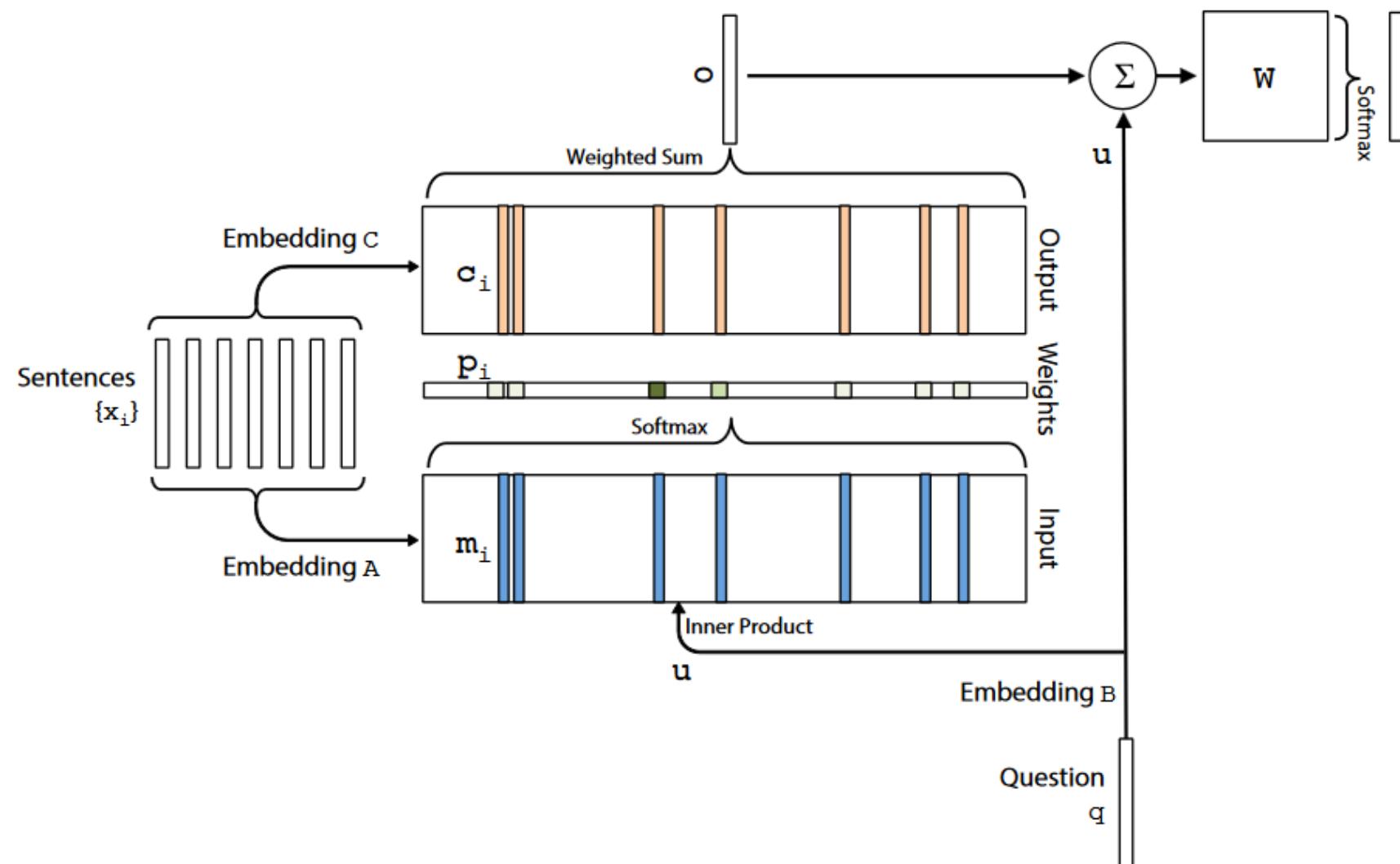
Julius is white.

Bernhard is green.

Q: What color is Brian?

A. White

Чатботы / экспертивные системы: механизмы памяти



«End-To-End Memory Networks» [Sukhbaatar S. и др., 2015 <https://arxiv.org/abs/1503.08895>]

Результат 2019 – GPT2

<https://blog.openai.com/better-language-models/>

[https://d4mucfpksywv.cloudfront.net/better-language-models/language models are unsupervised multitask learners.pdf](https://d4mucfpksywv.cloudfront.net/better-language-models/language%20models%20are%20unsupervised%20multitask%20learners.pdf)

Задачи

- **question answering**
- **machine translation**
- **reading comprehension**
 - **summarization**

обучение – новый датасет «WebText»

~ 1 млн web-страниц / 45 млн ссылок / 8 млн. документов
ссылки с Reddit ≥ 3 кармы (т.е. отбором человека)
удалили Wiki ! (чтобы тестировать на других датасетах)

Результат 2019 – GPT2

В основе – Language modeling

$$p(x) = \prod_{i=1}^n p(s_n | s_1, \dots, s_{n-1})$$

**Современная оценка таких вероятностей:
self-attention architectures ~ Transformer (Vaswani et al., 2017)**

p(output | input, task)

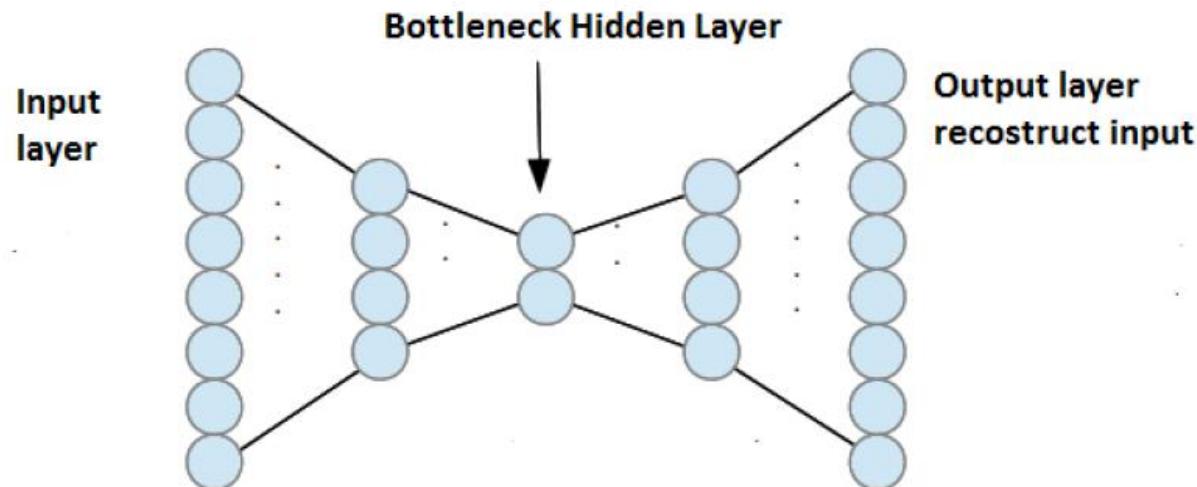
«переведи ...»

«ответь на вопрос ...»

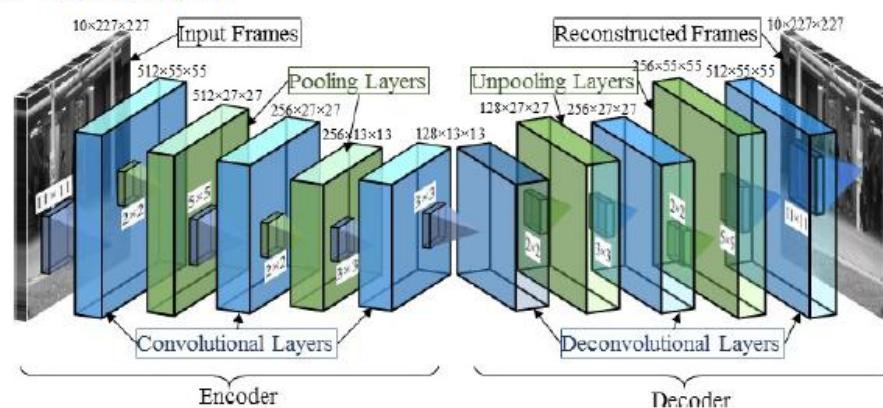
«TL;DR:»

Автокодировщики (Auto-encoders)

- ▶ Fully Contented Auto-encoder

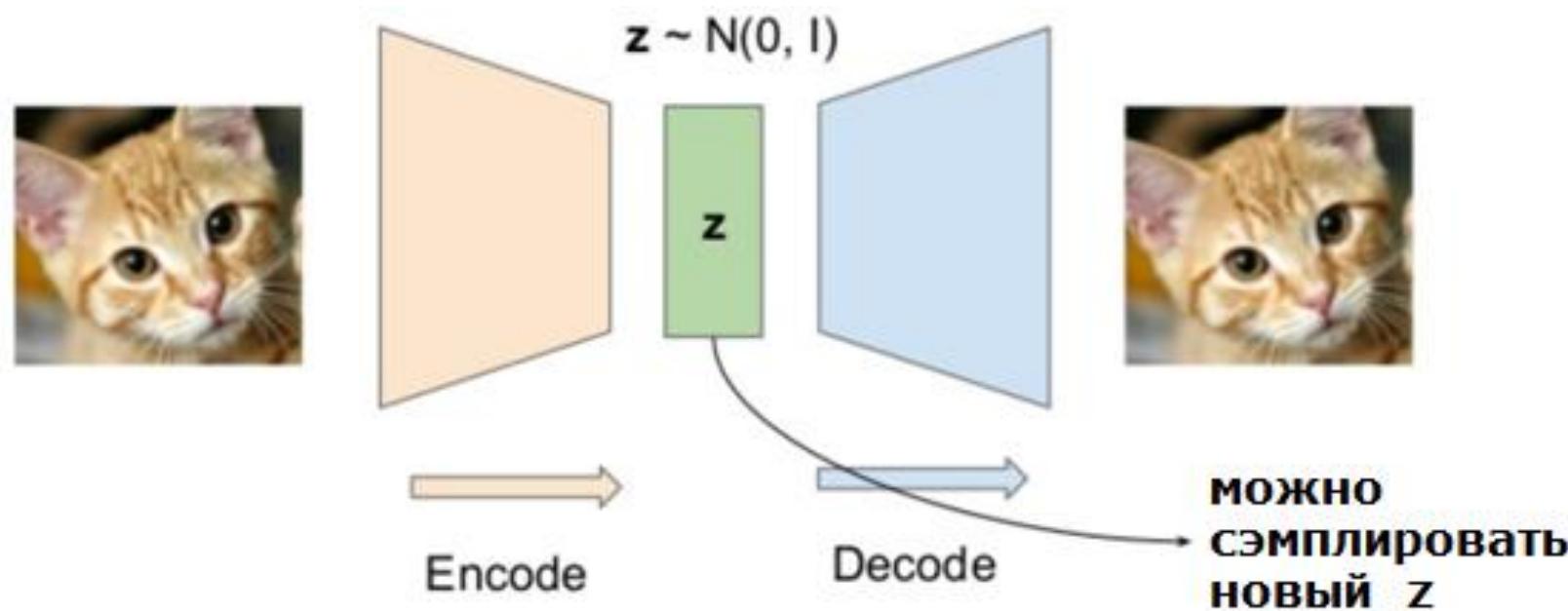


- ▶ Convolution Auto-encoder



VAE

**Обучать автокодировщик так, чтобы скрытые переменные
~ какое-то распределение**



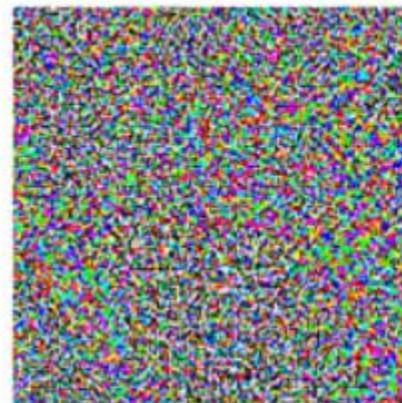
Состязательные атаки (Adversarial attacks)

Fast Gradient Sign Method, FGSM



x
“panda”
57.7% confidence

+ .007 ×



$\text{sign}(\nabla_x J(\theta, x, y))$
“nematode”
8.2% confidence

=



$x + \epsilon \text{sign}(\nabla_x J(\theta, x, y))$
“gibbon”
99.3 % confidence

Зашита:
обучение на атакованных данных
усложнение модели (ансамбли)
дистилляция (упрощение) сетей

Состязательные примеры – в тексте

Task: Spam filtering. **Classifier:** LSTM. **Original label:** 100% Spam. **New label:** 89% Non-Spam.

Text: your application **petition** has been **aceepted** **recognized** thank you for your **loan** **borrower** **request** **petition** , which we **recieved** yesterday , your **refinance** **subprime** **application** **petition** has been **aceepted** **recognized** good credit or not , we are ready to give you a \$ oov loan , after further review , our lenders have established the lowest monthly payments . approval process will take only 1 minute . please visit the confirmation link below and fill-out our short 30 second secure web-form . http : oov

Task: Sentiment analysis. **Classifier:** CNN. **Original label:** 81% Positive. **New label:** 100% Negative.

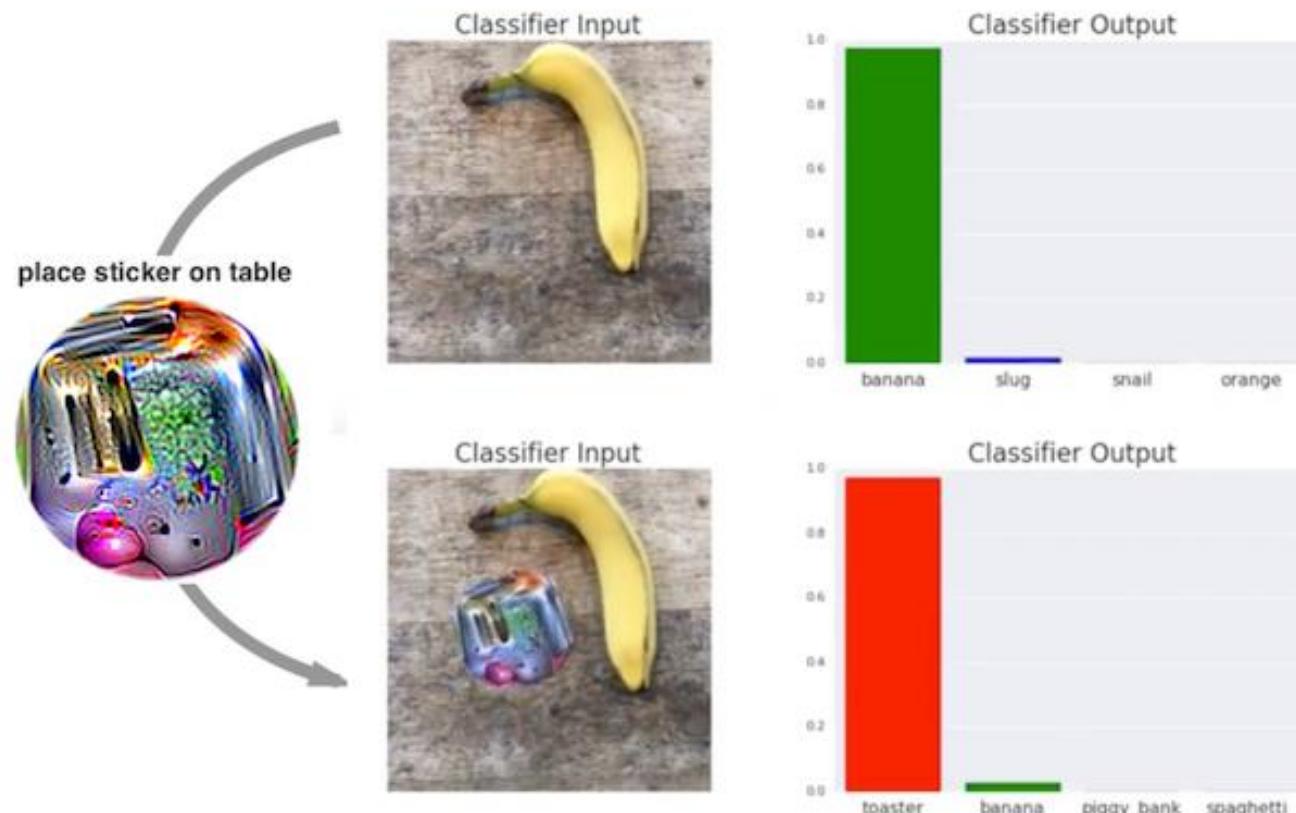
Text: i **went** **moved** to wing wednesday which is all-you-can-eat wings for \$ oov even though they raise the prices it 's **still** **ever** really great deal . you can eat as many wings you want to get all the different **flavors** **tastes** and have a good time enjoying the atmosphere . the girls are smoking hot ! all the types of **sauces** **dressings** are awesome ! and i had at least 25 wings in one sitting . i would **definitely** **certainly** go again just **simply** not every **wednesday** **friday** maybe once a month .

Task: Fake news detection. **Classifier:** Naive Bayes. **Original label:** 97% Fake. **New label:** 100% Real

Text: trump supporter whose **brutal** **ferocious** beating by black **mob** **gangsta** was caught on **video** **tape** **asks** **demands** : “ what happened to america ? ” [video] , ” david oov , a 49 year **old** **former** **chicago** **rochester** man who was brutally beaten by a **mob** **lowlife** of black democrats **asks** **demands** , “ what happened to america ? ” here is his very **sad** **disappointing** story

<https://openreview.net/pdf?id=r1QZ3zbAZ>

Состязательные участки – Adversarial patch



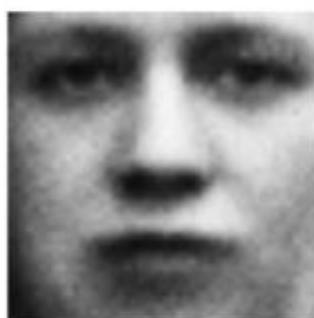
**Пример внедряемого участка,
который отвечает за нужную классификацию**

Brown, T. B., Mané, D., Roy, A., Abadi, M., & Gilmer, J. (2017). Adversarial Patch, (Nips).
<http://arxiv.org/abs/1712.09665>

На эту же тему: DeepFakes



Генерация изображений



2014



2015



2016



2017



2018

<https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf>

http://research.nvidia.com/publication/2017-10_Progressive-Growing-of

Generative Adversarial Networks [Goodfellow et al., 2014]

Генератор – сеть, которая порождает объект (изображение) из шума

Дискриминатор – сеть, различающая настоящие и сгенерированные объекты

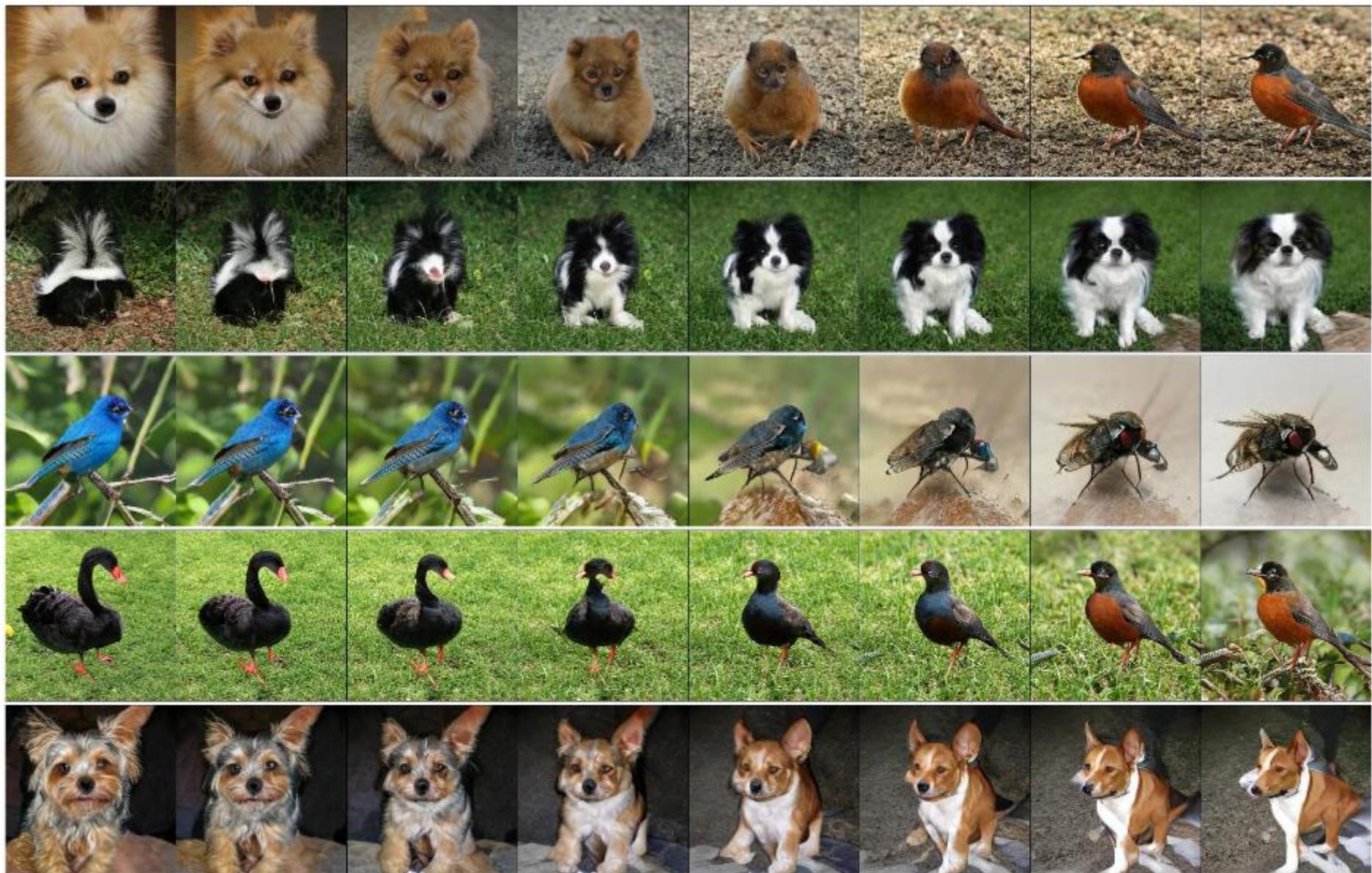
Самое главное:

**По сути, дискриминатор – дифференцируемая функция ошибки!
Эту идею можно использовать там, где нет подходящих функций ошибок...**

Задачи

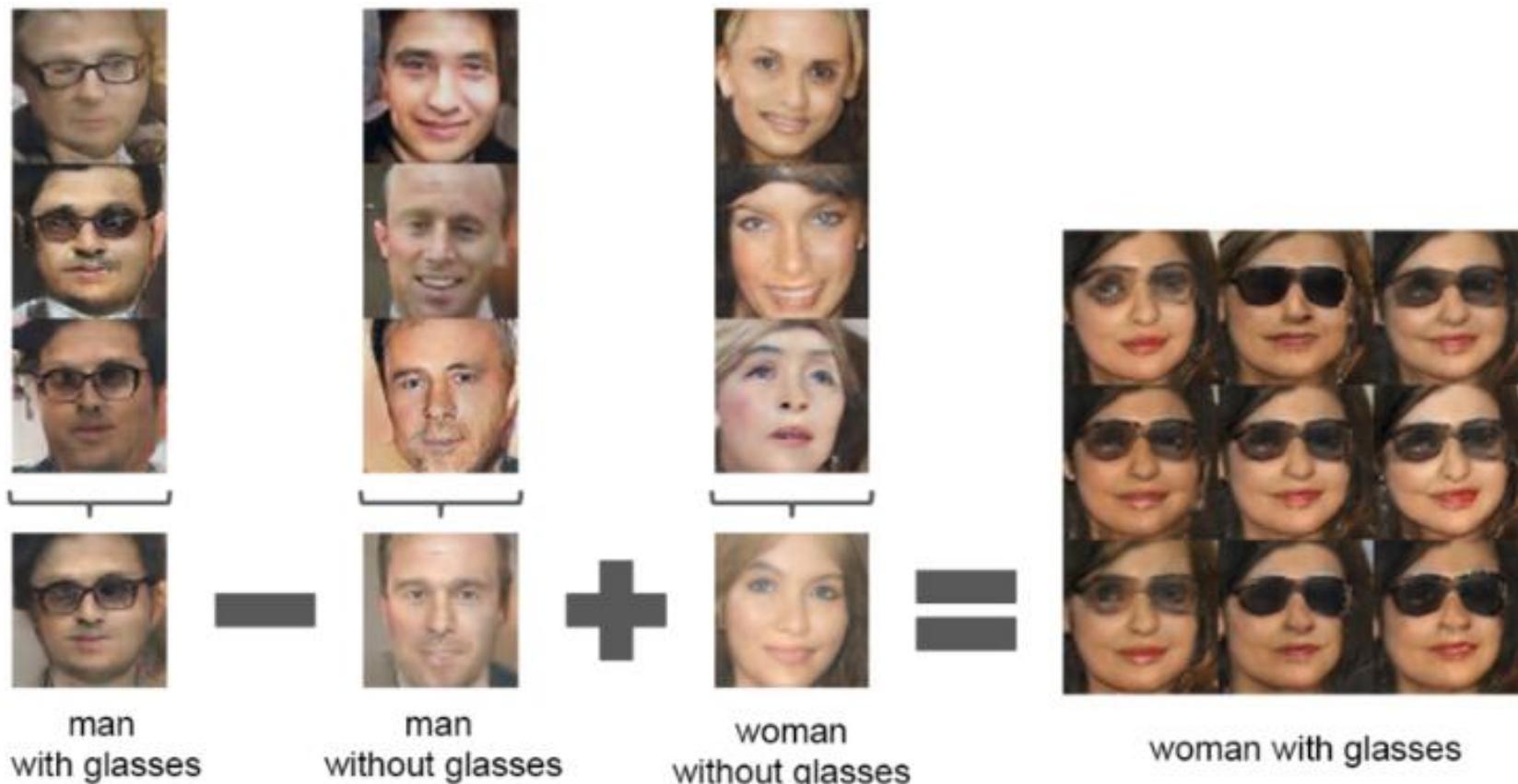
- **улучшение изображений (Image Inpainting)**
- **улучшение звука (speech enhancement)**
- **генерация изображений (Image Generation)**
- **супер-разрешение (Super-resolution)**

BigGAN: Генерация изображений / интерполяция



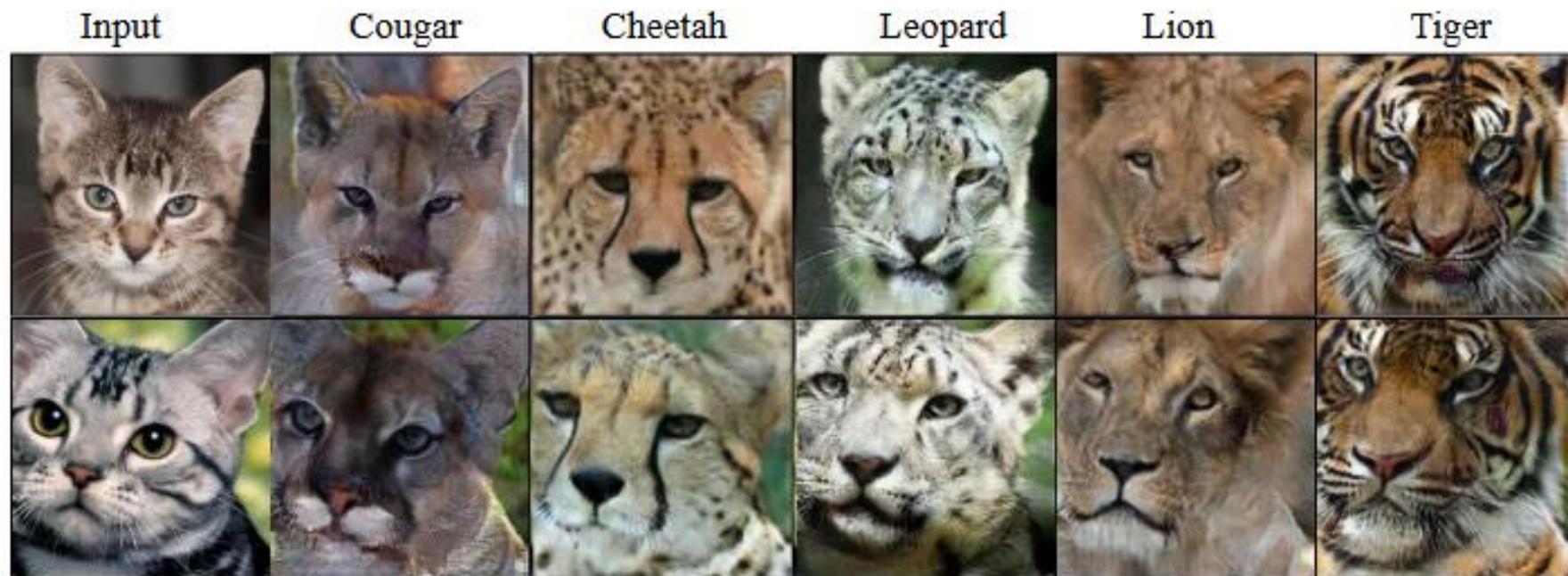
<https://arxiv.org/pdf/1809.11096.pdf> (учим одновременно и обратную)

DCGAN



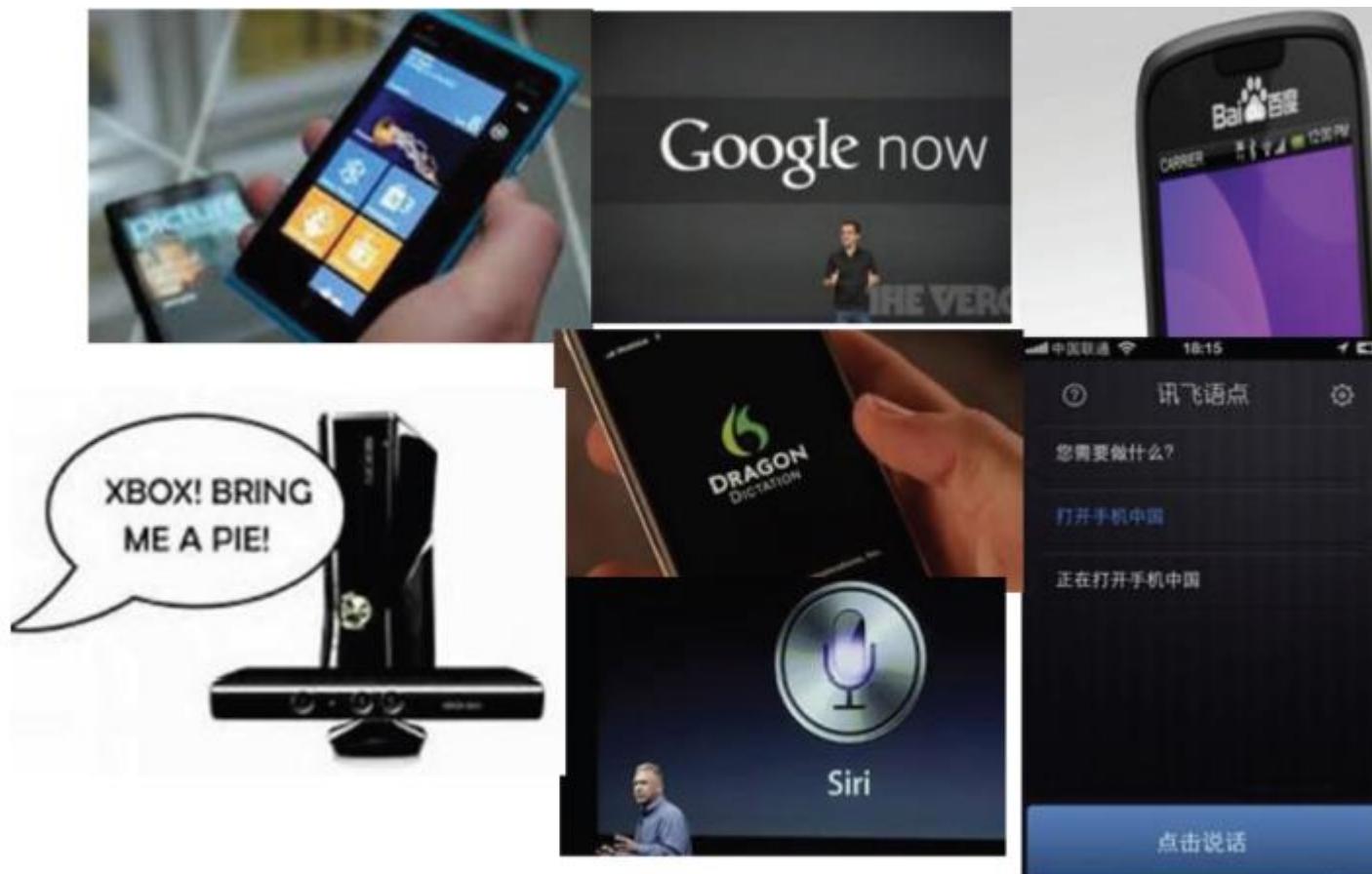
[Radford et al, ICLR 2016]

Трансляция изображений (Image-to-Image Translation)



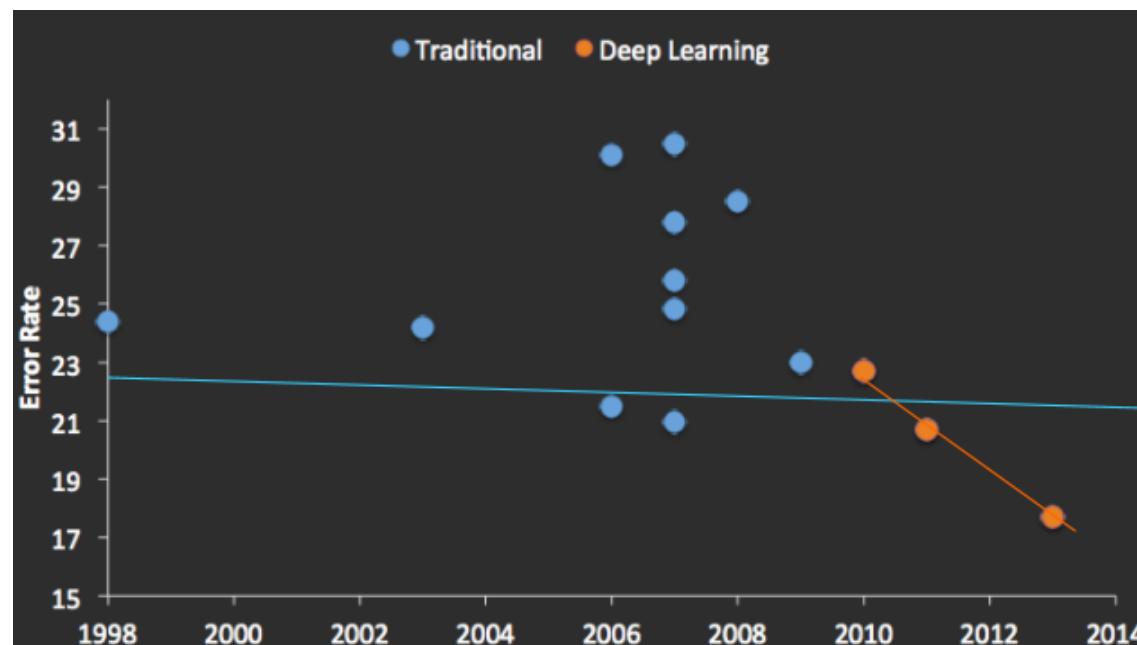
<https://arxiv.org/pdf/1703.00848.pdf>

Автоматическое распознавание речи: ASR

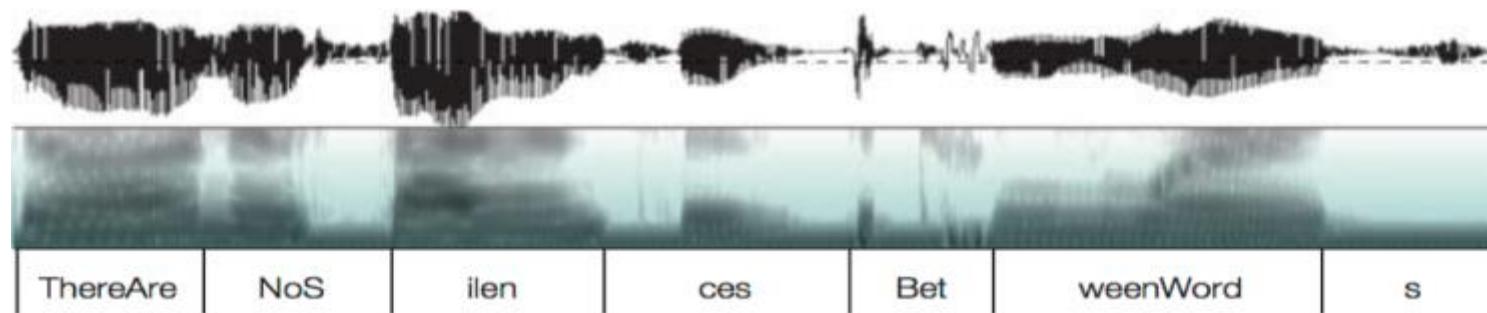


слайды Антон Бахтин (Facebook Research)

Автоматическое распознавание речи: ASR



<https://en.wikipedia.org/wiki/TIMIT>



[Kuhl et al. 2004]

Решения в распознавании речи

Мир коммерческих решений

- Google
- Amazon Alexa
- IBM Watson
 - Siri
- Yandex

Свободные системы

- CMU Sphinx
- Kaldi
- Mozilla DeepSpeech

Kaldi

инструменты для распознавания речи (больше для исследования)

C++, лицензия Apache v2.0

гибкий понятный код для расширения и изменения!

BLAS и LAPACK (поддержка линейной алгебры)

DeepSpeech

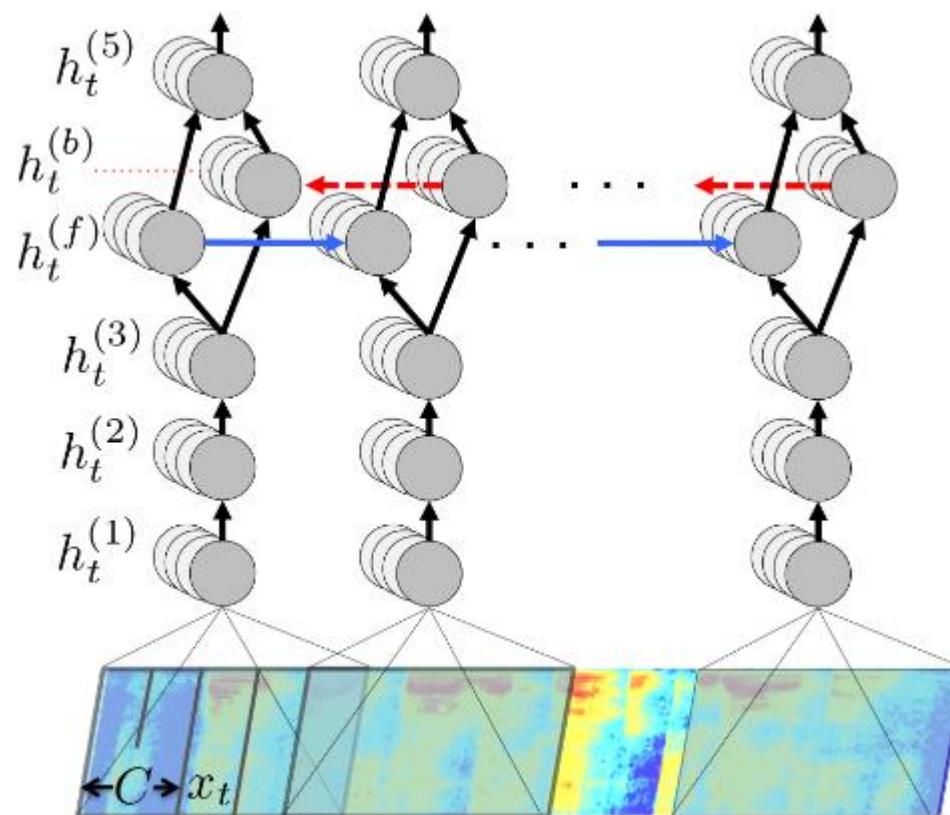
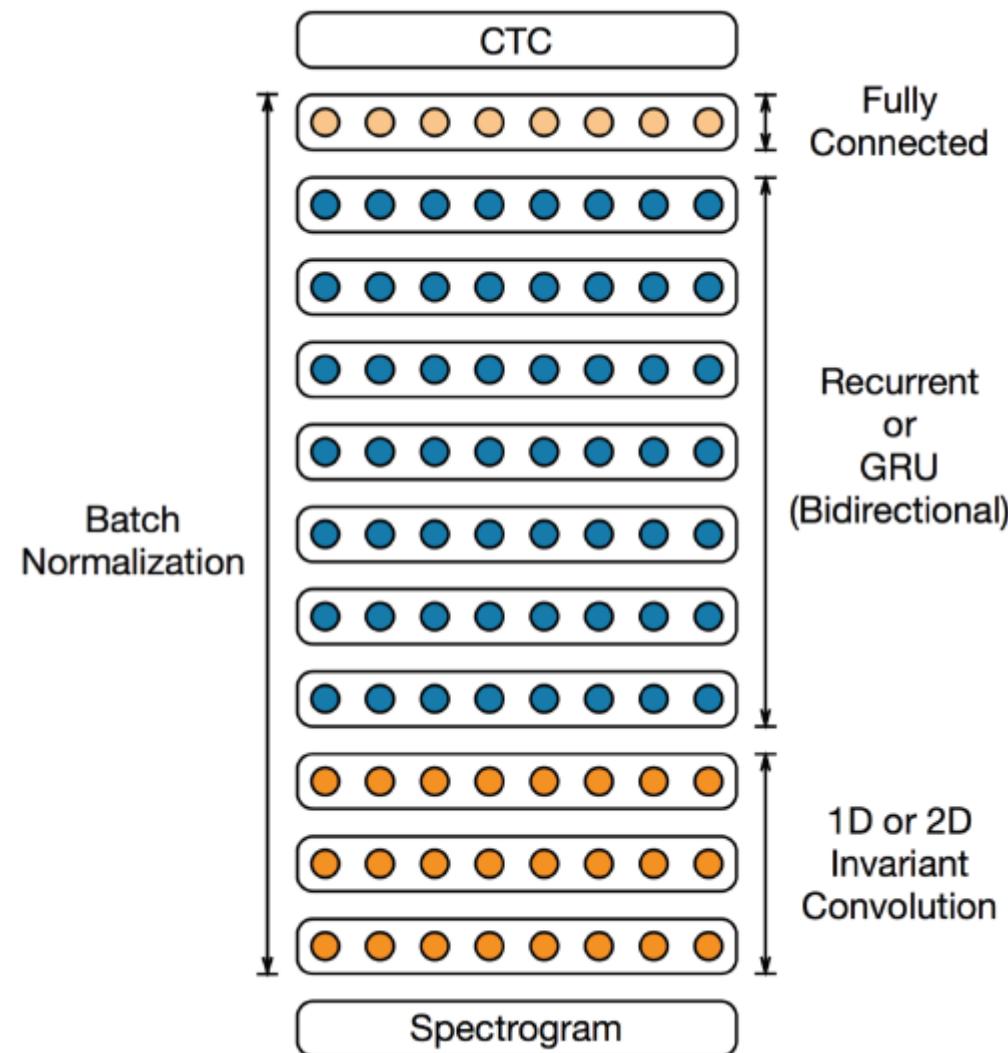


Figure 1: Structure of our RNN model and notation.

**RNN 5 слоёв = 3 обычных (clipped ReLu) + 1 двусторонний
рекуррентный + 1 обычный (на вход сумма выходов 4го)
выход = softmax для алфавита**

DeepSpeech2

Архитектура только «разрослась», суть DS сохранилась...

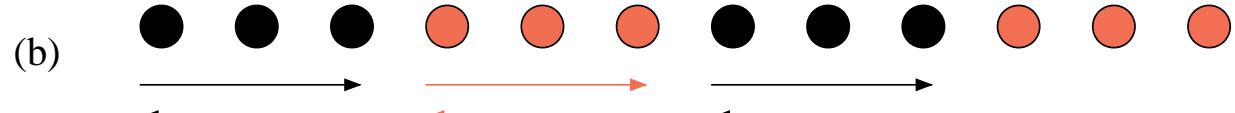


Latency Constrained Bidirectional GRU

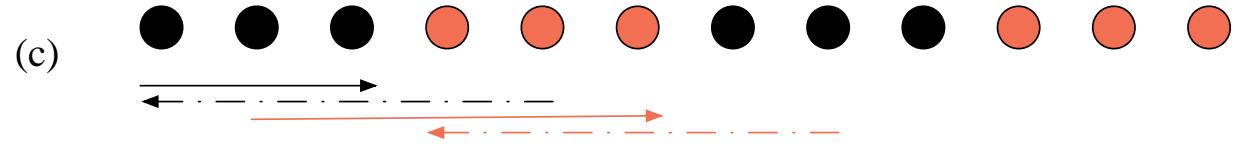
Bidirectional RNN



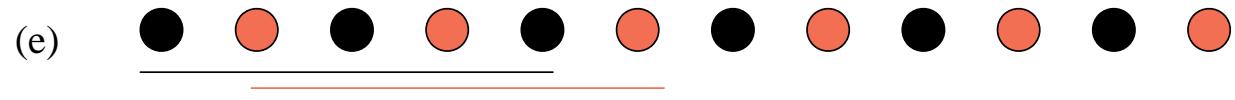
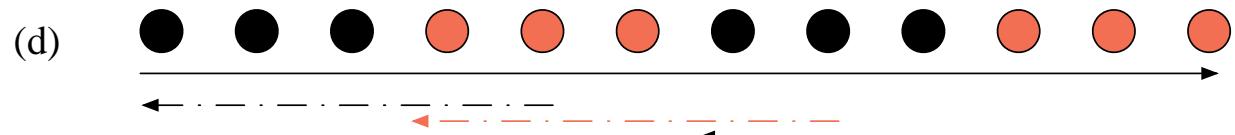
Chunked RNN



Chunked RNN with overlapping



LC-BGRU layer



Look-ahead Convolution

+ модель языка! (Cold Fusion, Shallow Fusion, ...)

Задачи в обработке аудио text-to-speech, TTS (Генерация речи по тексту)

WaveNet [van den Oord и др., 2016]

<https://deepmind.com/blog/wavenet-generative-model-raw-audio/>

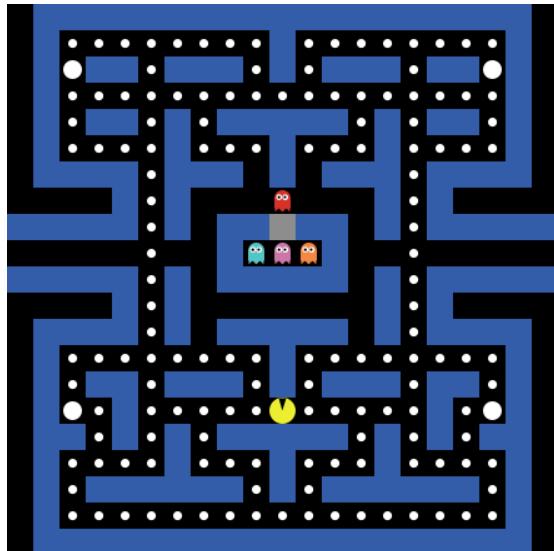
<https://arxiv.org/pdf/1609.03499.pdf>

Tacotron 2

- произношение
 - громкость
 - скорость
 - интонация

<https://research.googleblog.com/2017/12/tacotron-2-generating-human-like-speech.html>

Задачи ИИ в играх Atari, шахматы, го, дота



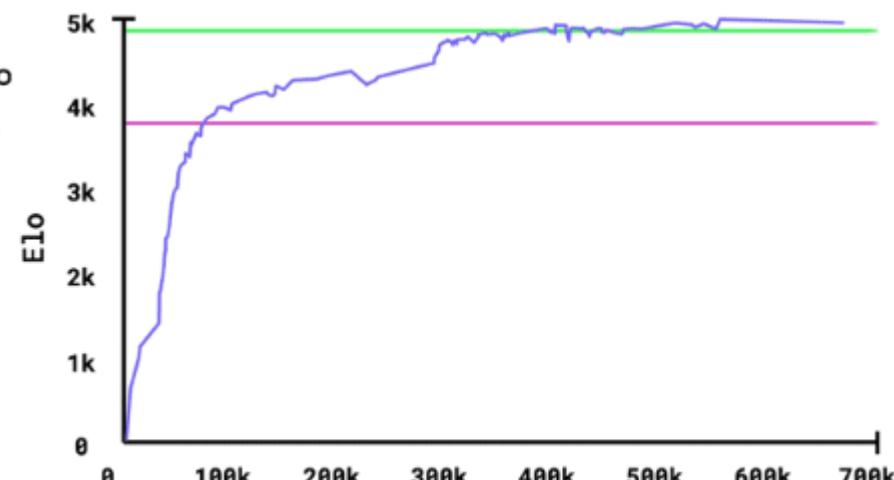
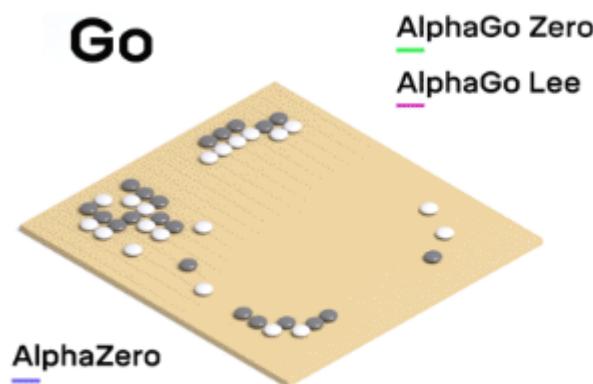
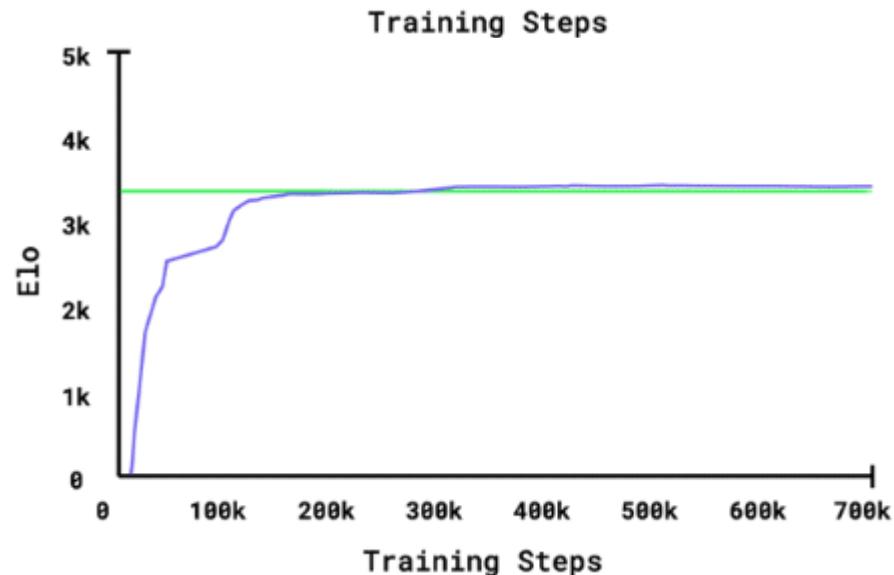
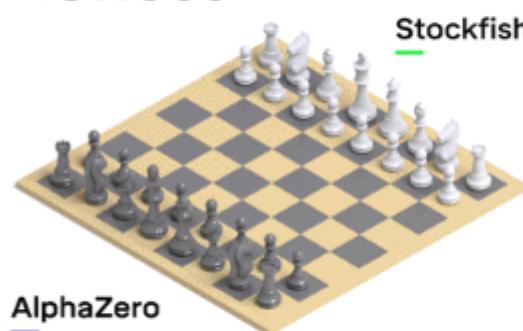
2013, тоже Google DeepMind
Обучение играть в классические игры 70-80-ых
для игровой консоли Atari 2600 (49 видеоигр, в 22
лучше человека).
НС ничего не знает про правила игры!
Использует изображение на входе и информацию
об очках!



9-15.03.2016
AlphaGo (Google DeepMind) — Ли Седоль
4 – 1

1920 процессоров и 280 графических
процессоров

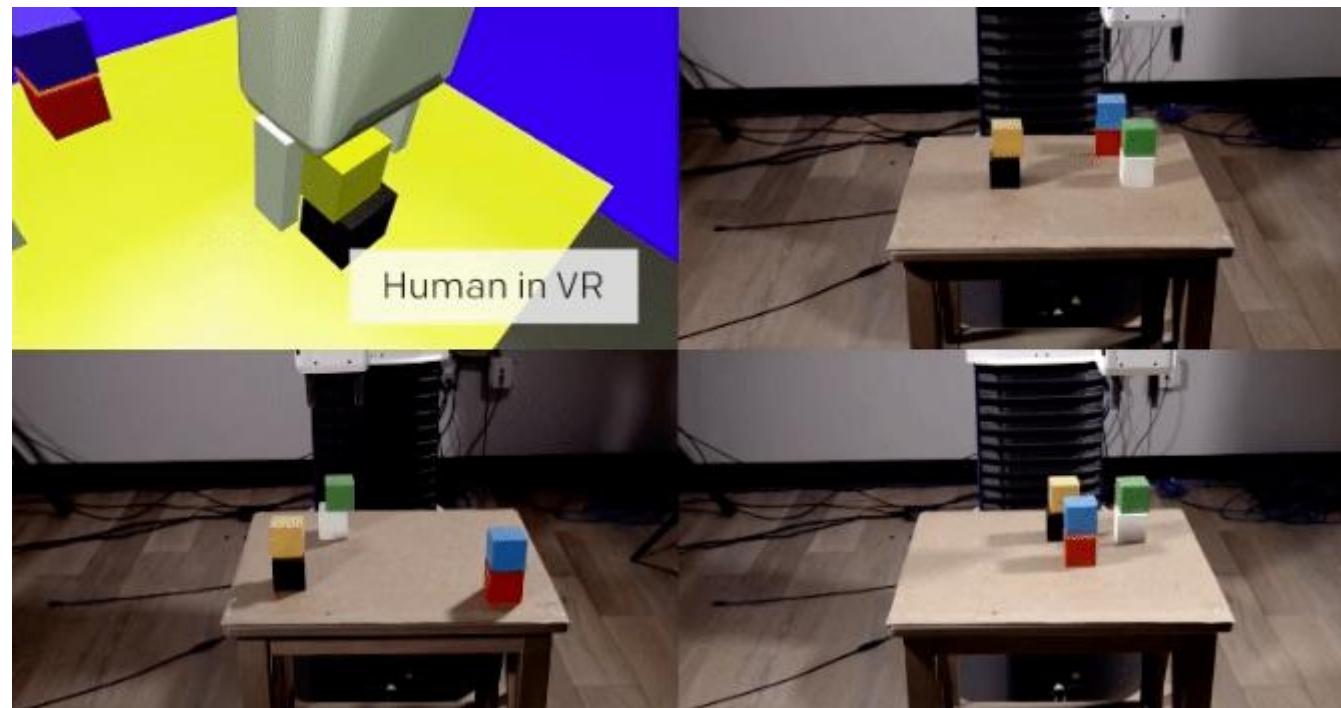
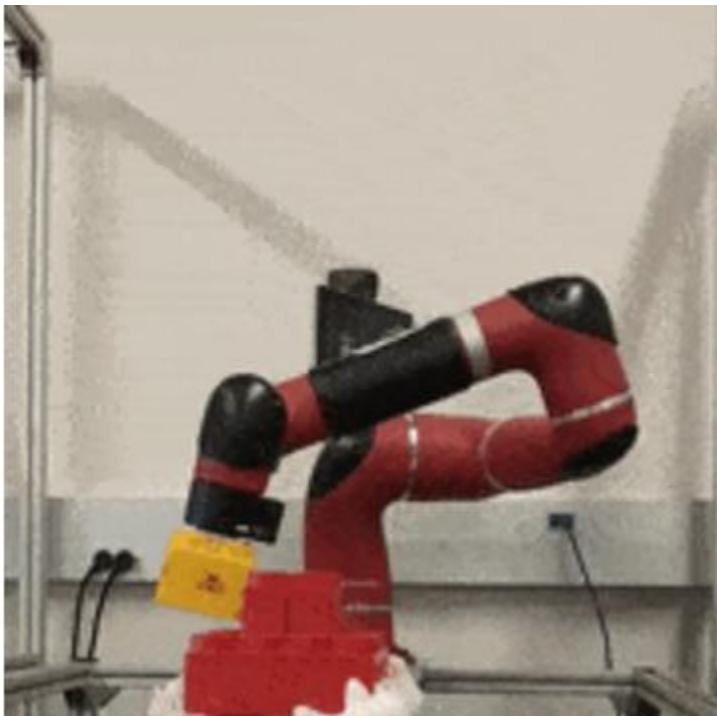
Задачи ИИ в играх

Go**Chess**

Alpha Go Zero: Играет без анализа человеческого опыта

<https://arxiv.org/abs/1712.01815>

Автомобили-беспилотники, роботы-манипуляторы



<http://bair.berkeley.edu/blog/2017/10/06/soft-q-learning/>

<https://blog.openai.com/robots-that-learn/>

RL: Многорукие бандиты



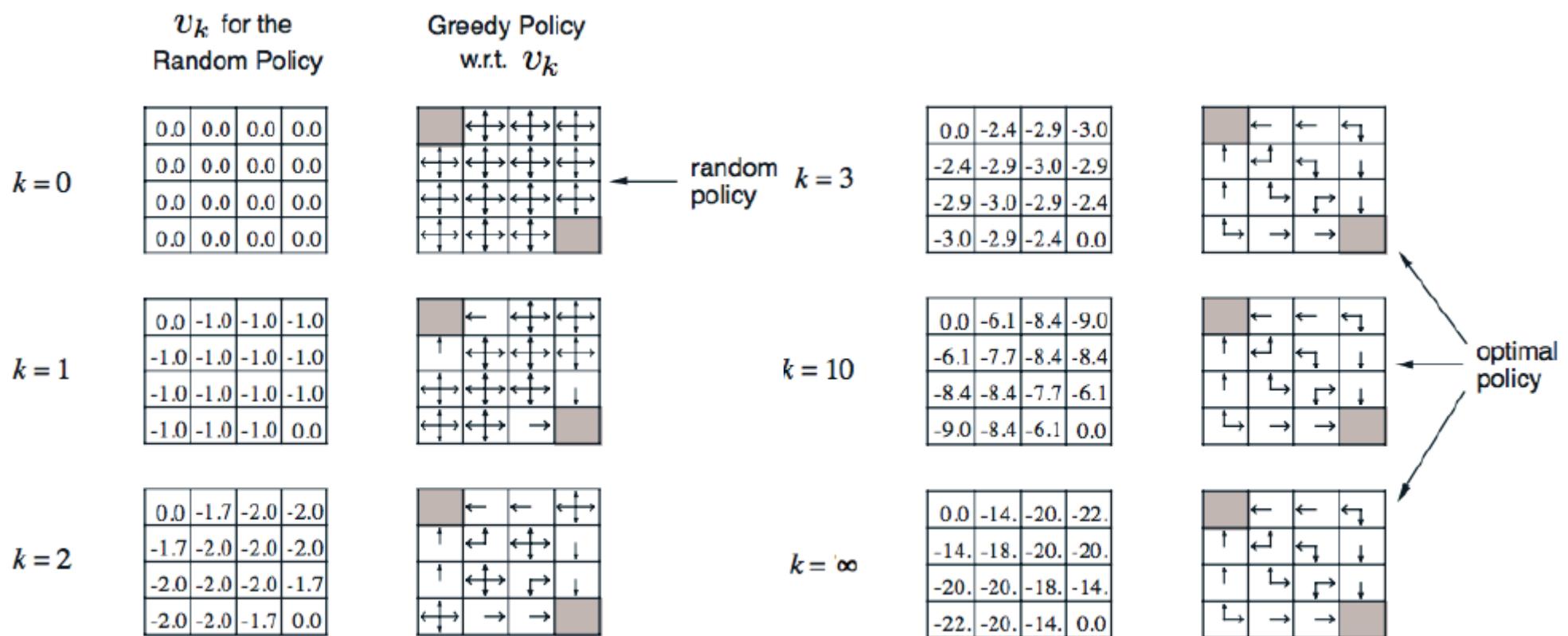
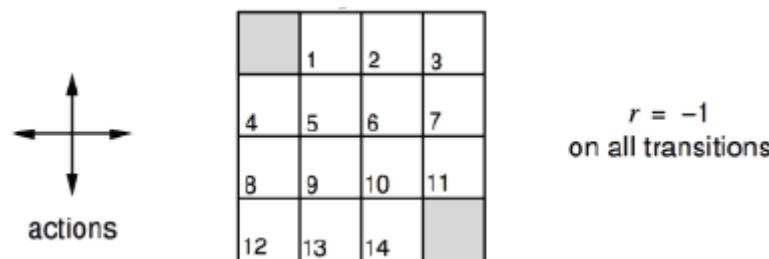
UCB1

$n_a = |\{i : a_i = a\}|$ – сколько раз выбирали a

$Q(a) = \frac{1}{n_a} \sum_{a_i=a} r_i$ – доход от решения a

$$a_t = \arg \max_a \left(Q(a) + \alpha \sqrt{\frac{2 \ln t}{n_a}} \right)$$

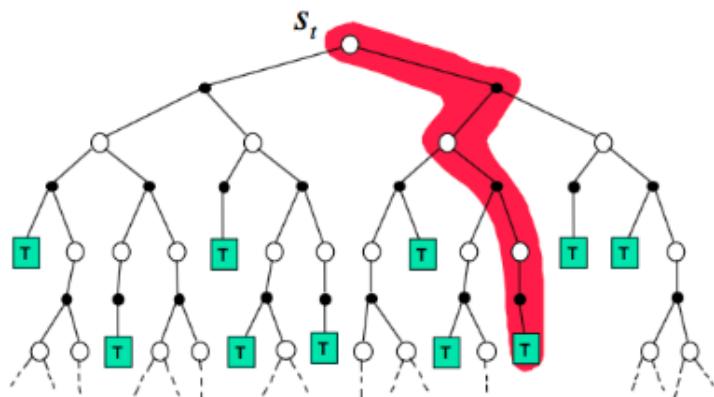
RL: Policy Iteration



RL

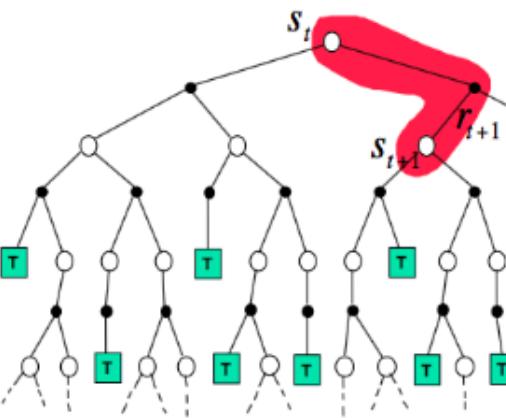
Monte-Carlo

$$V(S_t) \leftarrow V(S_t) + \alpha (G_t - V(S_t))$$



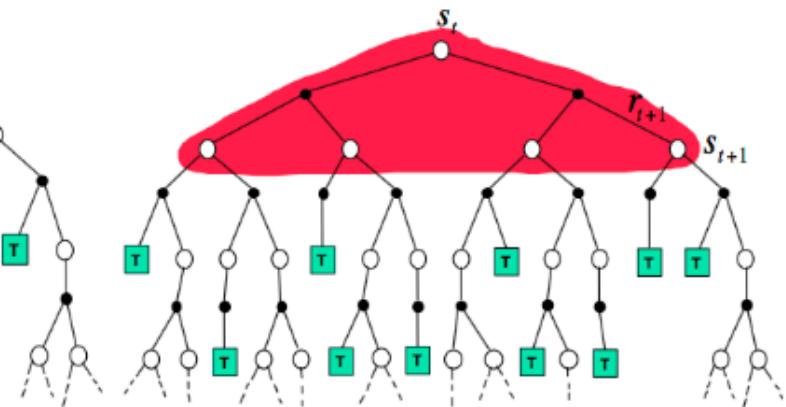
Temporal-Difference

$$V(S_t) \leftarrow V(S_t) + \alpha (R_{t+1} + \gamma V(S_{t+1}) - V(S_t))$$



Dynamic Programming

$$V(S_t) \leftarrow \mathbb{E}_\pi [R_{t+1} + \gamma V(S_{t+1})]$$



Ссылки

Материалы курса «Глубокое обучение»

<https://github.com/Dyakonov/DL>

Документация

<https://www.tensorflow.org/tutorials/>

<https://pytorch.org/tutorials/>

Лучшая книга по глубокому обучению

Ian Goodfellow, Yoshua Bengio, Aaron Courville «Deep Learning» An MIT Press book <http://www.deeplearningbook.org/>

Другие курсы

Классика DL CS231n:

«Convolutional Neural Networks for Visual Recognition from Stanford»

<http://cs231n.stanford.edu/syllabus.html>