

**курс «Глубокое обучение»**

# **Введение – общий обзор**

**Александр Дьяконов**

**7 февраля 2022 года**

## Глубокое обучение (Deep Learning)

**курс для студентов 3 курса ММП ВМК МГУ – читается с весны 2019 года**  
**каждый год материал расширяется**  
**часть курса читается в OzonMasters**

**лекции со слайдами**  
**семинары ~ pytorch + практикум**

**понедельник, 12:05, онлайн**  
**Отчётность: проекты + экзамен**

**Страница курса:** <https://github.com/Dyakonov/DL/>

**Общение:** телеграм-чат (не пишите в личку)

<https://classroom.google.com>

## Лектор

**Дьяконов Александр Геннадьевич**

### Другие курсы лектора

- «**Введение в машинное обучение**» бакалавриат ВМК (спецкурс)
- «**Методы машинного обучения**» 3 поток 3 курс ВМК (лекции)
- «**МОАД**» бакалавриат 3 поток 4 курс ВМК (лекции)
- «**ПЗАД**» Магистратура 1г. ММП ВМК (лекции + семинары)

## История AI / DL

**1943 – McCulloch, Pitts – нейрон (суперпозиция линейной и пороговой функций)**

**1950 – 70 – «коннективизм», модель нейрона**

**1969 – Minsky, Papert – книга «Персептроны»**

**1970 – 80 – зима ИИ**

**1980 – Fukushima – неокогнитрон (neocognitron), вид CNN**

**1980 – 1990 – экспертные системы, обратное распространение**

**1985 – Hinton, Sejnowski – машины Больцмана**

**1990 – 2000 – зима**

**1997 – Schmidhuber, Hochreiter – LSTM**

**1998 – LeCunn – LeNet (CNNs) распознавание символов**

**2003 – Bengio – NLM (neural language modeling)**

**2009 – распознавание речи**

**2012 – AlexNet – победа на ImageNet ConvNets (Krizhevsky и др.)**

**2012 – Google – DL ASR**

**2014 – информационный / инвестиционный бум**

**2014 – NMT seq2seq, VAE, GAN**

**2015 - ResNet**

**2016 – 17 – DeepMind – Alphago-Lee Sedol, AlphaZero**

**2018 – BERT(Google), Style GANs, VQ-VAE**

**2019 – Super-human performance on GLUE**

## Особенности DL

**Классическое МО**

генерация признаков → модель

генерация признаков → нелинейности (kernel),  
преобразования (PCA), селекция → модель

**Глубокое обучение**

модель

**Глубокое обучение (Deep learning) – обучение многослойных моделей / графов вычислений**

## AI vs человек: когда компьютер превзошёл человека...

<b>2011</b>	<b>Jeopardy!</b>	<b>IBM Watson</b>
<b>2015</b>	<b>Atari Games</b>	<b>Google DeepMind (в большинстве игр)</b>
<b>2015</b>	<b>ImageNet</b>	<b>ResNet</b>
<b>2016-17</b>	<b>Go</b>	<b>AlphaGo, AlphaGo Master, AlphaGo Zero // Google DeepMind</b>
<b>2017</b>	<b>Poker</b>	<b>Libratus // CMU DeepStack // University of Alberta</b>
<b>2017</b>	<b>Ms. Pac-Man</b>	<b>Maluuba // Microsoft</b>
<b>2018</b>	<b>Chinese - English Translation</b>	<b>Microsoft</b>
<b>2018</b>	<b>Capture the Flag (Quake III Arena)</b>	<b>DeepMind</b>
<b>2018</b>	<b>DOTA 2 (с ограничениями!)</b>	<b>OpenAI с любительскими командами</b>
<b>2018</b>	<b>Starcraft II</b>	<b>Alphastar // DeepMind</b>
<b>2019</b>	<b>Detect diabetic retinopathy (DR) with specialist-level accuracy</b>	<b>Google AI</b>

## DL: первые успехи

## Первый прорыв – речь

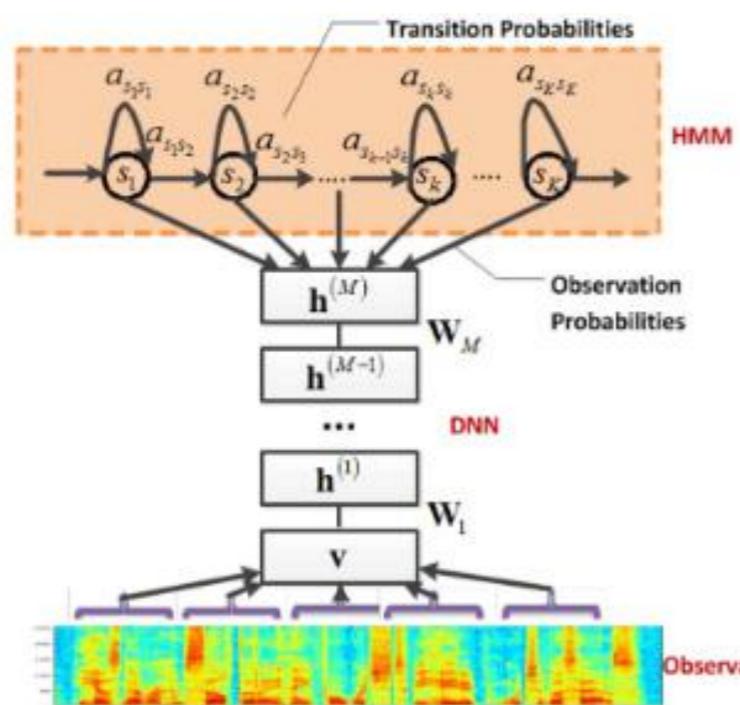
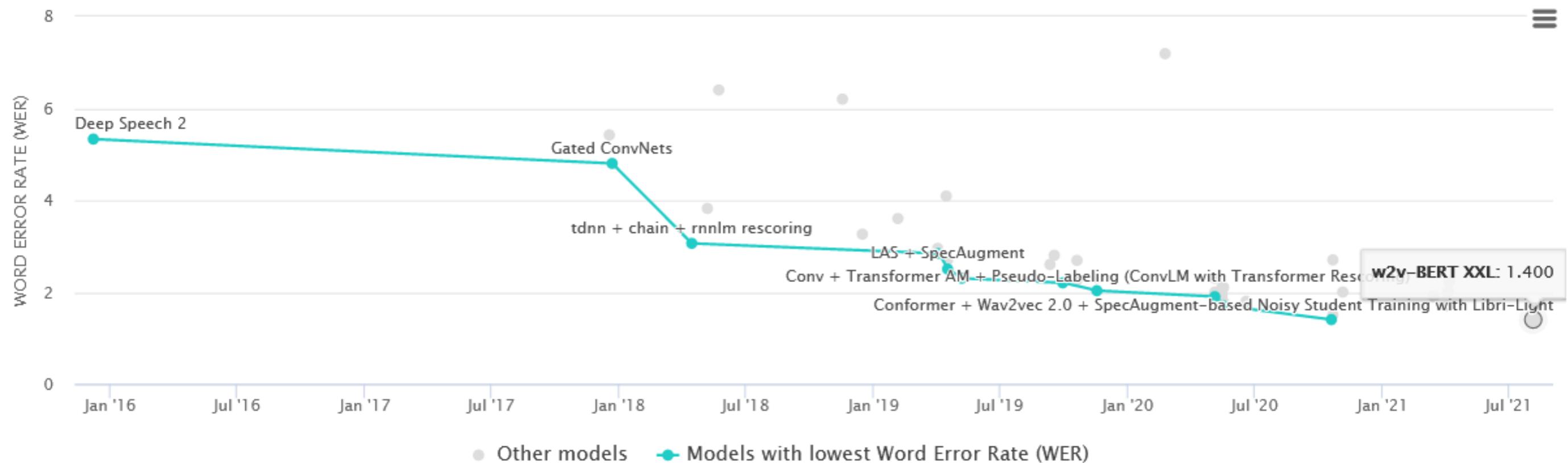


Fig. 1. Diagram of our hybrid architecture employing a deep neural network. The HMM models the sequential property of the speech signal, and the DNN models the scaled observation likelihood of all the senones (tied tri-phone states). The same DNN is replicated over different points in time.

Acoustic model Measuring WER	RT03S FSH	Hub5 SWB
Traditional (GMM) (2012)	<b>27.4</b>	<b>23.6</b>
Deep Learning (Dahl et al. 2012)	<b>18.5</b> (-33%)	<b>16.1</b> (-32%)
Xiong et al. (2017)		<b>5.8</b>

Dahl et al. «Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition» (2010)  
<https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/dbn4lvcslr-transaslp.pdf>

## Распознавание речи



<https://paperswithcode.com/sota/speech-recognition-on-librispeech-test-clean>

Есть также устаревшее <https://www.eff.org/ai/metrics>

## DL: первые успехи

## Соревнование по распознаванию дорожных знаков CNN (2011)

## Results

## IJCNN2011 COMPETITION

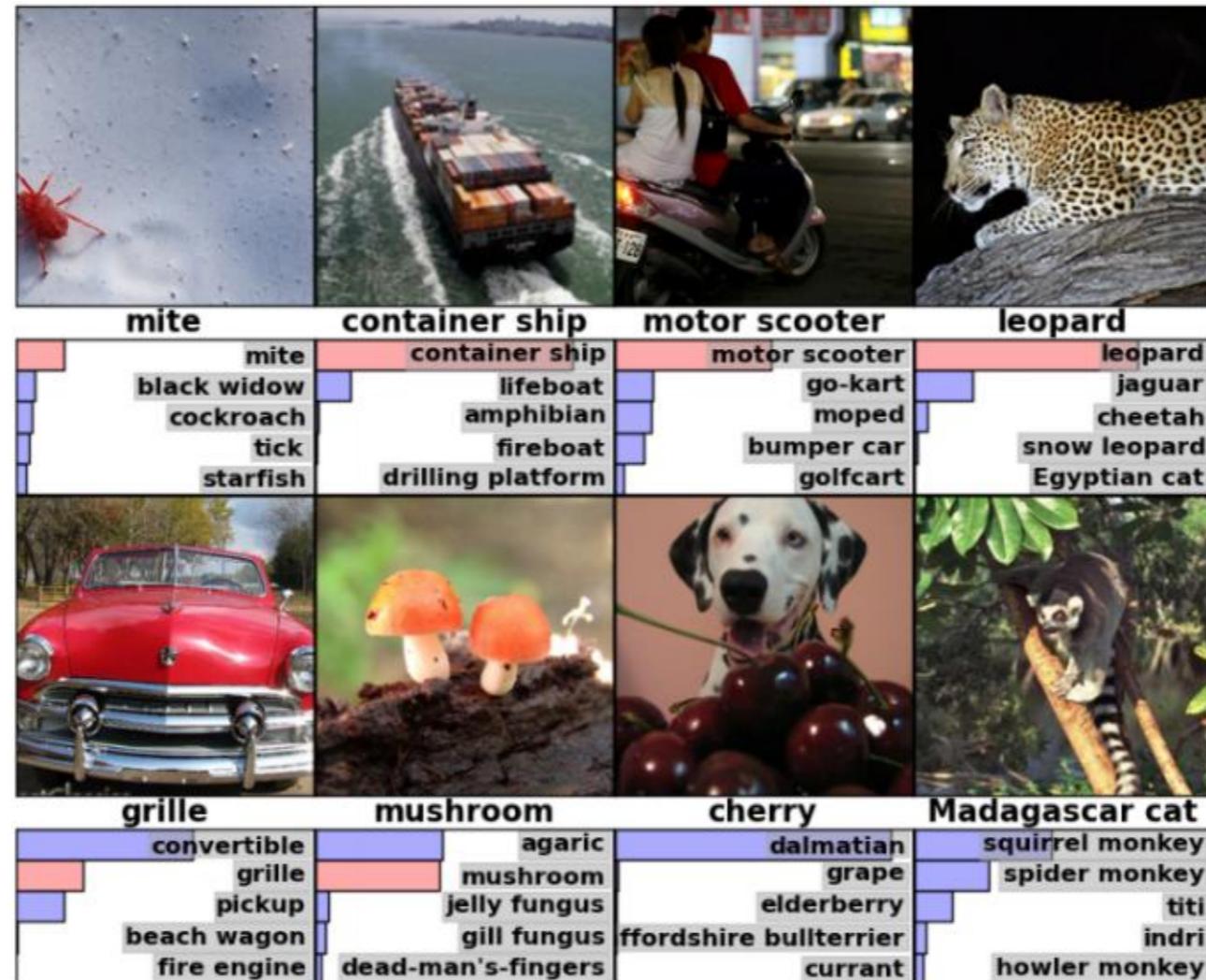
TEAM	METHOD	TOTAL	SUBSET ALL SIGNS
[3] IDSIA ★	Committee of CNNs	99.46%	99.46%
[1] INI-RTCV ★	Human Performance	98.84%	98.84%
[4] sermanet ★	Multi-Scale CNNs	98.31%	98.31%
[2] CAOR ★	Random Forests	96.14%	96.14%
[6] INI-RTCV	LDA on HOG 2	95.68%	95.68%
[5] INI-RTCV	LDA on HOG 1	93.18%	93.18%
[7] INI-RTCV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



Превзошёл точность человека!

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Классификация изображений



Krizhevsky, A., et al. ImageNet Classification with Deep Convolutional Neural Networks, NIPS, 2012

## Революция в машинном обучении



# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Пример Классификации: QuickDraw



<https://quickdraw.withgoogle.com/>  
<https://arxiv.org/pdf/1704.03477.pdf>

## Сервисы по классификации

Определение породы собаки по фотографии / марки авто / товара



**Shiba Inu**

Lively, independent-minded, aloof  
toward strangers



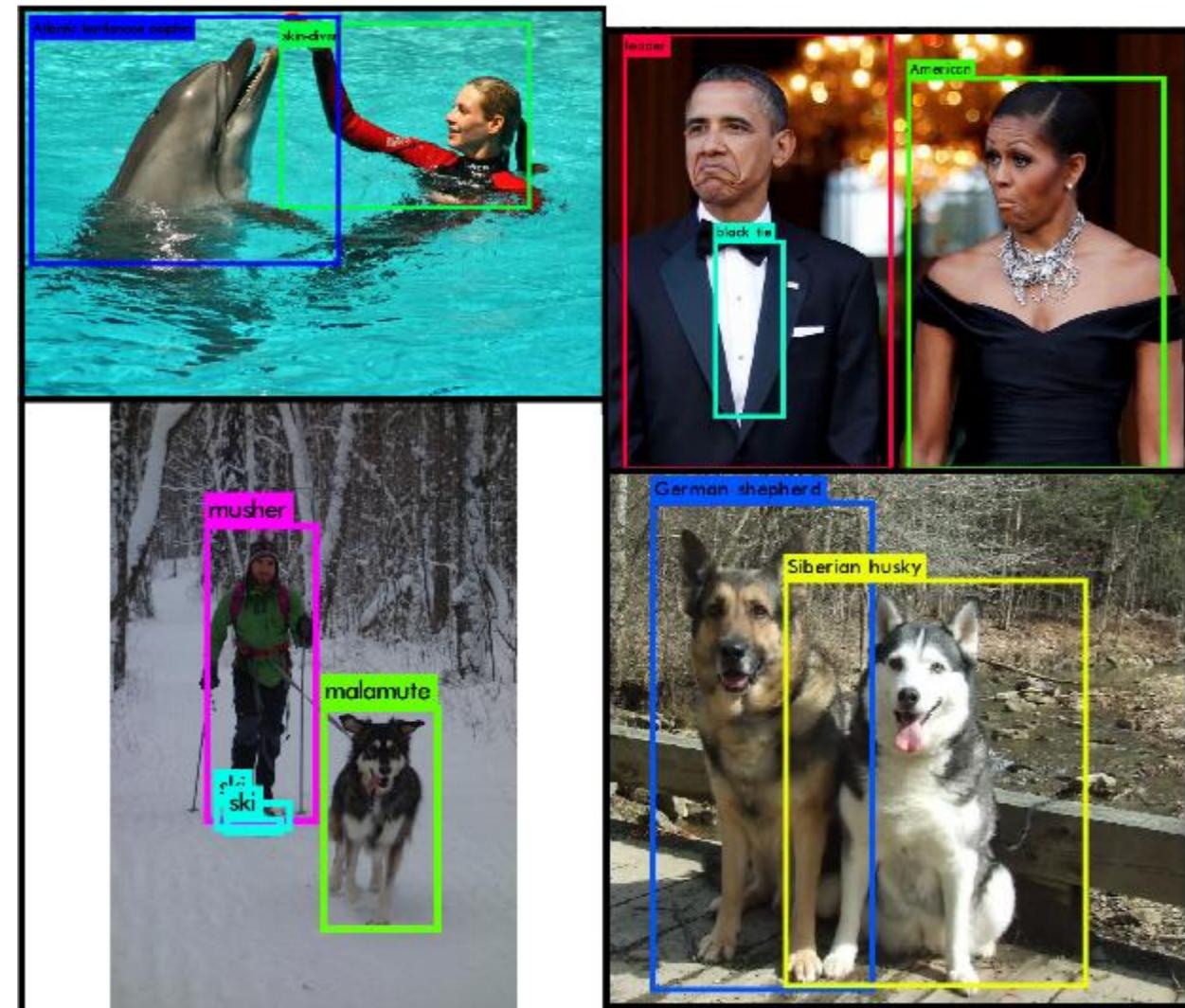
**German Shepherd Dog**

Must be kept busy, intense, highly  
intelligent

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

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Обнаружение объектов (Object Detection)



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

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Сегментация объектов



Не и др., 2012

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

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Удаление фона (Background removal)

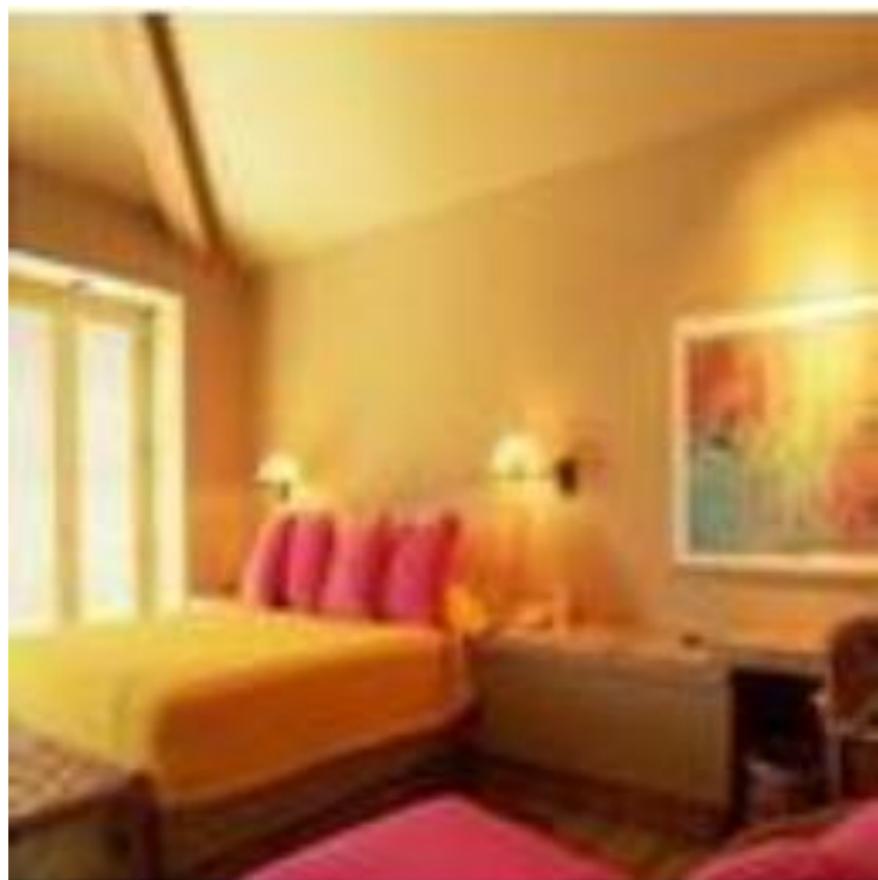


<https://towardsdatascience.com/background-removal-with-deep-learning-c4f2104b3157>

<https://github.com/YunanWu2168/Background-removal-using-deep-learning>

## Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

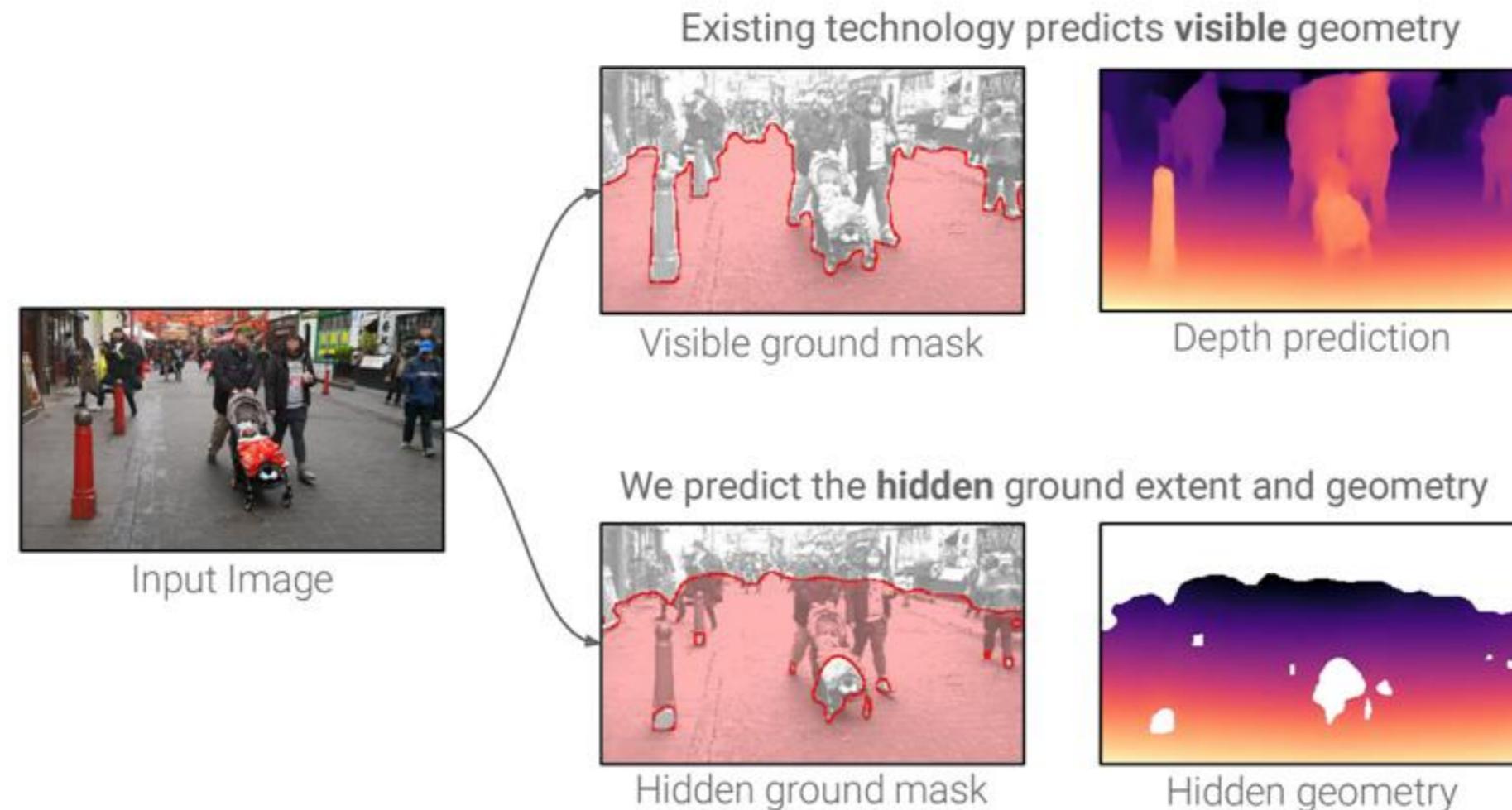
### Распознавание сцен (Scene recognition)



Bedroom

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Предсказание геометрии пространства



**По одной фотографии – где здесь могут находиться объекты**

<https://github.com/nianticlabs/footprints>

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

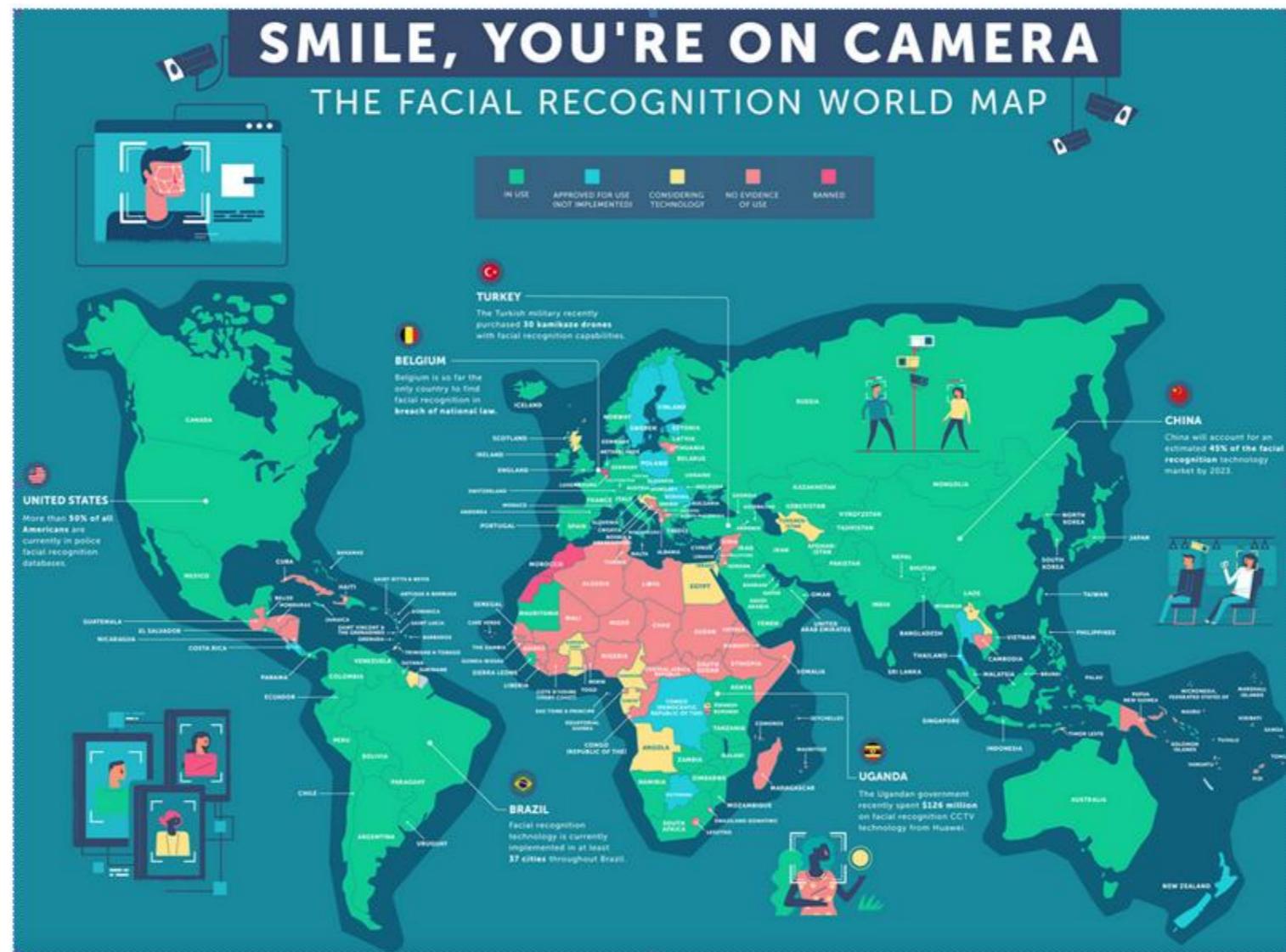
## Определение позы (Human pose estimation)



S. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. CoRR, abs/1602.00134, 2016.

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

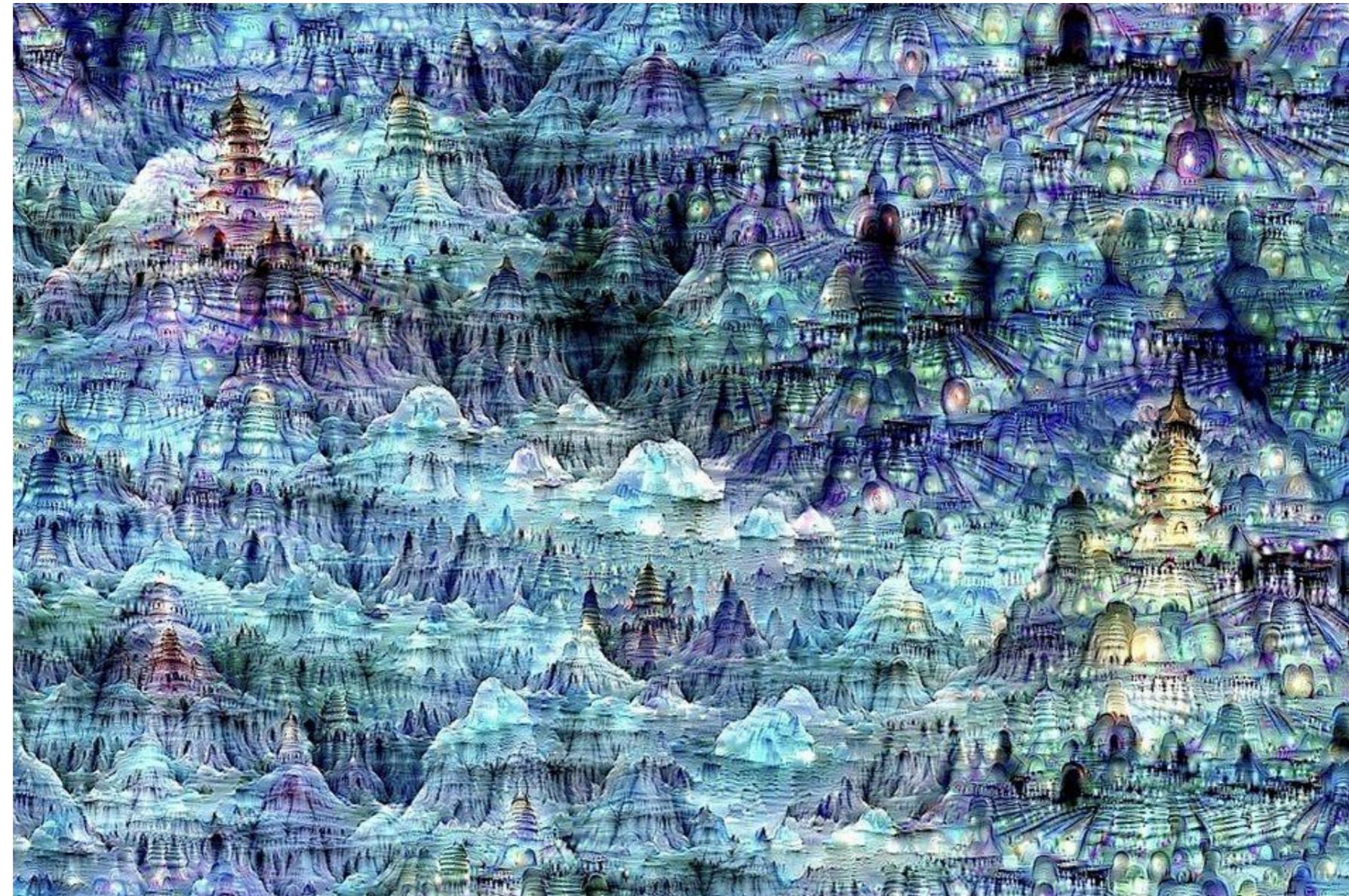
## Распознавание лиц



<https://www.visualcapitalist.com/wp-content/uploads/2020/05/Facial-Recognition-World-Map-Full-Size.html>

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Генерация изображений



## Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

### Генерация фотoreалистичных изображений и лиц



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

[http://research.nvidia.com/publication/2017-10\\_Progressive-Growing-of](http://research.nvidia.com/publication/2017-10_Progressive-Growing-of)

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

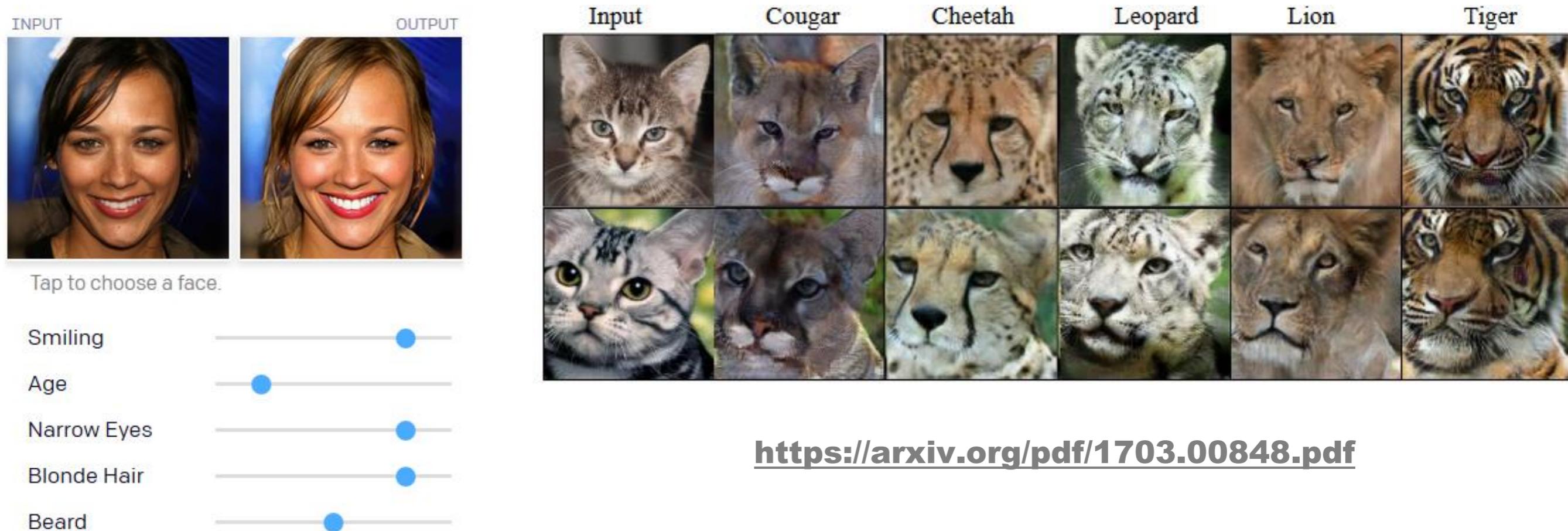
## Интерполяция в пространстве изображений



<https://arxiv.org/pdf/1809.11096.pdf>

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Контролируемое изменение изображений, трансляция изображений

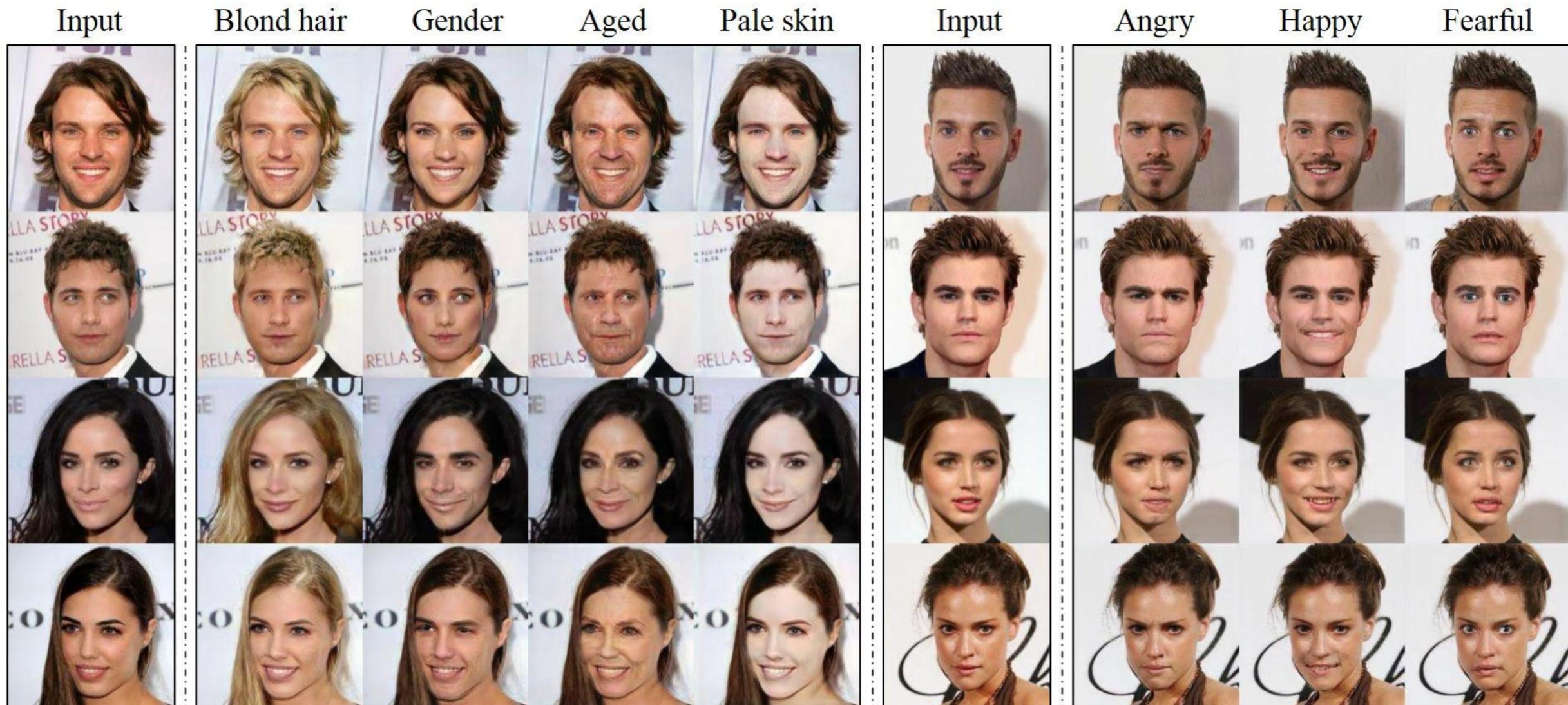


<https://arxiv.org/pdf/1703.00848.pdf>

<https://openai.com/blog/glow/>

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Контролируемое изменение изображений, трансляция изображений

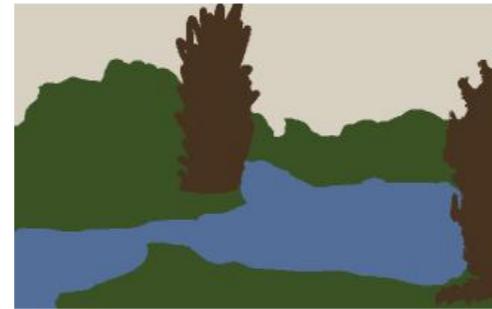


# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Генерация изображений по эскизам

Synthesized Image

#NeuralDoodle



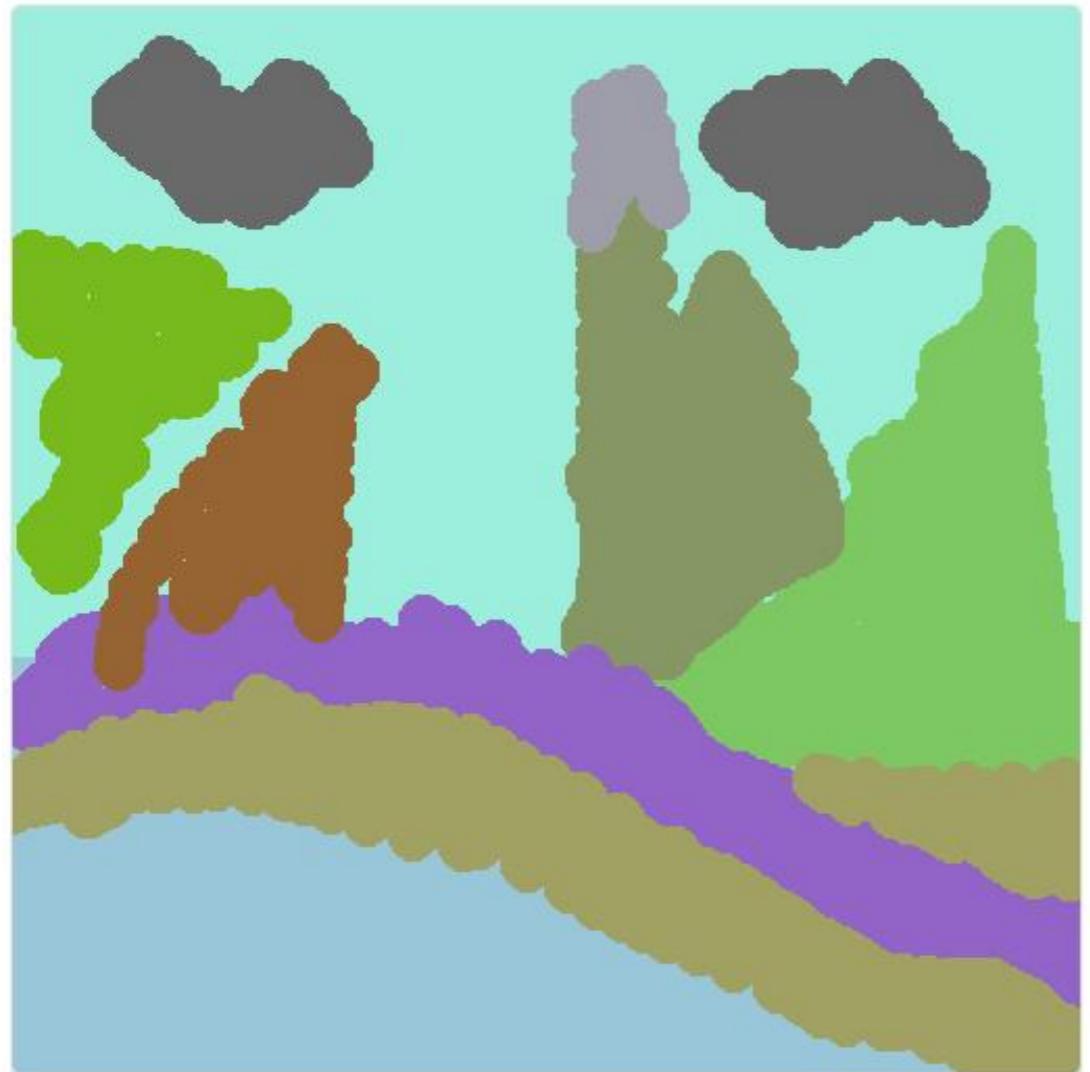
#NeuralDoodle



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

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Генерация изображений по эскизам



Clouds  
Fog  
Hill  
Mountain  
River  
Rock  
Sea  
Sky  
Snow  
**Stone**  
Water



**GauGAN:** <https://www.nvidia.com/en-us/research/ai-playground/>

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Улучшение изображений, дорисовка изображений (Image completion)



[http://hi.cs.waseda.ac.jp/~iizuka/projects/completion/data/completion\\_sig2017.pdf](http://hi.cs.waseda.ac.jp/~iizuka/projects/completion/data/completion_sig2017.pdf)

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

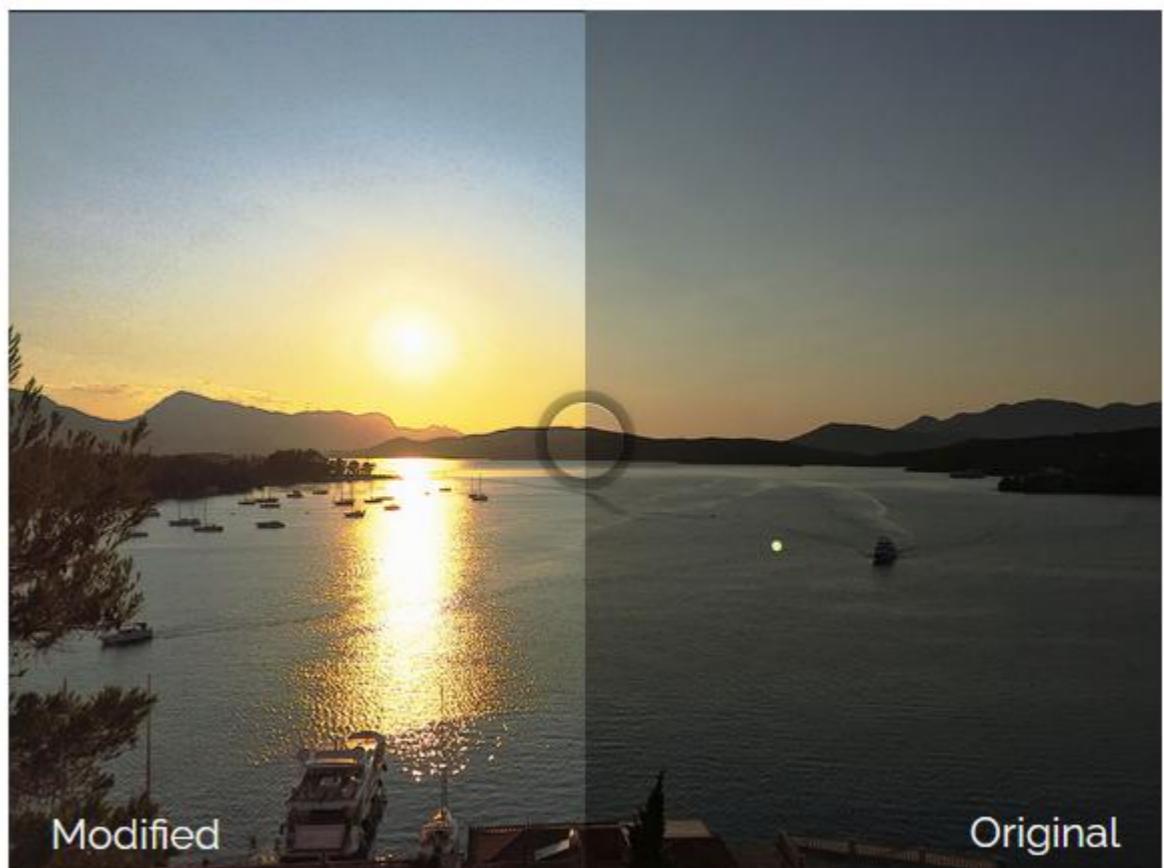
## Улучшение изображений, супер-разрешение (Super-Resolution)



<https://arxiv.org/abs/1609.04802>

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Улучшение фотографий



**в обучении не пары (плохое хорошее), а примеры хороших и что надо улучшить**  
<http://www.vision.ee.ethz.ch/~ihnatova/wespe.html>

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Стилизация изображений (перенос стиля)



<https://deepart.io/>

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



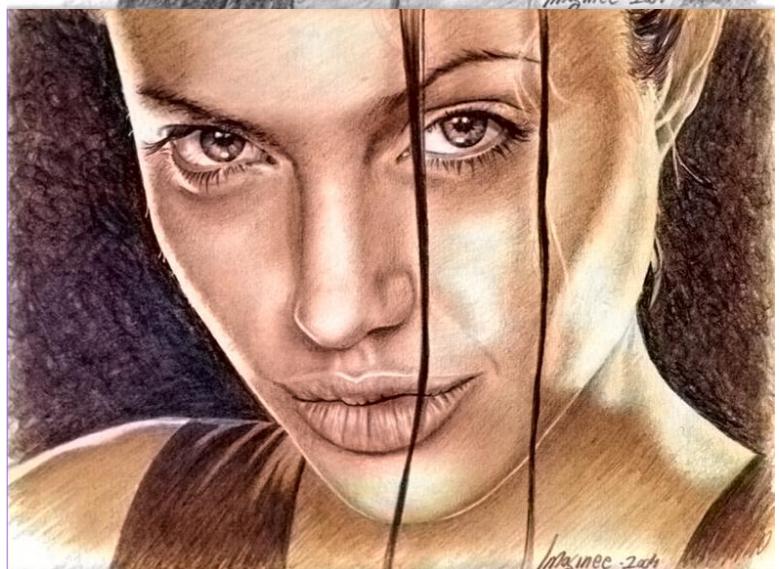
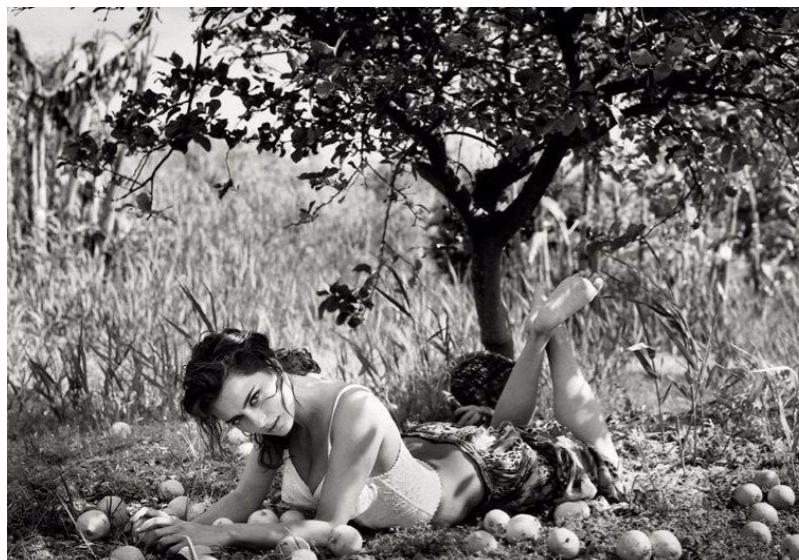
(a) Input image

(b) Reference style image

(e) Our result

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

## Раскрашивание изображений



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

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

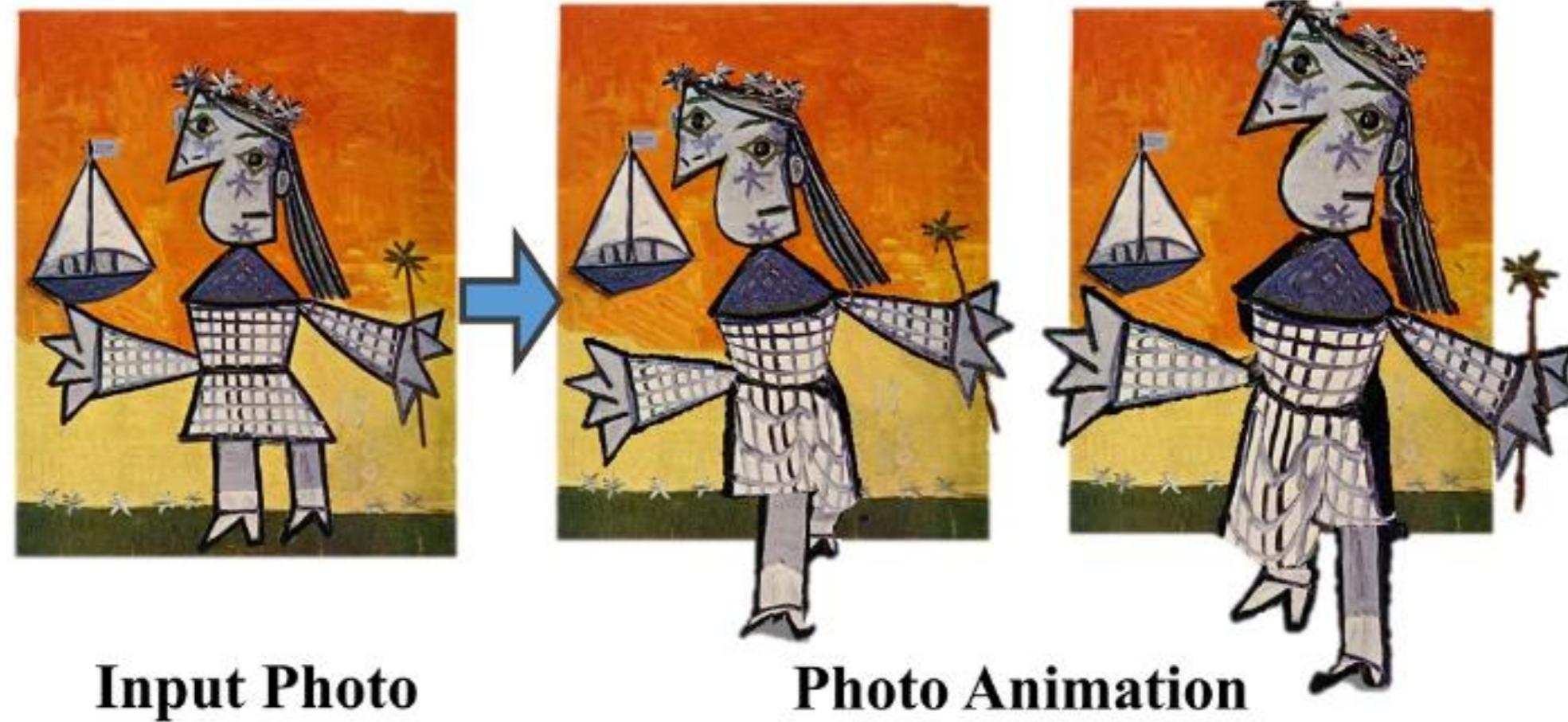
## Фото по изображению



<https://neural.love/portraits>

# Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

Обработка изображений: image → video



<https://arxiv.org/pdf/1812.02246.pdf>

## Реконструкция изображений / панорам



«LayoutNet: Reconstructing the 3D Room Layout from a Single RGB Image», 2018 // <https://arxiv.org/abs/1803.08999>

## Дизайн / креатив

### Генерация объектов / представление объектов



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

## Понимание изображений

### 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/>

A person riding a motorcycle on a dirt road.



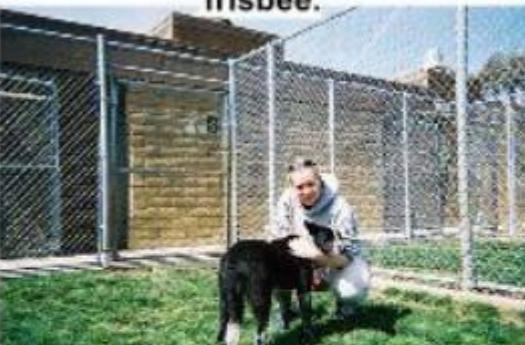
Two dogs play in the grass.



A skateboarder does a trick on a ramp.



A dog is jumping to catch a frisbee.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.



A little girl in a pink hat is blowing bubbles.



A refrigerator filled with lots of food and drinks.



A herd of elephants walking across a dry grass field.



A close up of a cat laying on a couch.



A red motorcycle parked on the side of the road.



A yellow school bus parked in a parking lot.



Describes without errors

Describes with minor errors

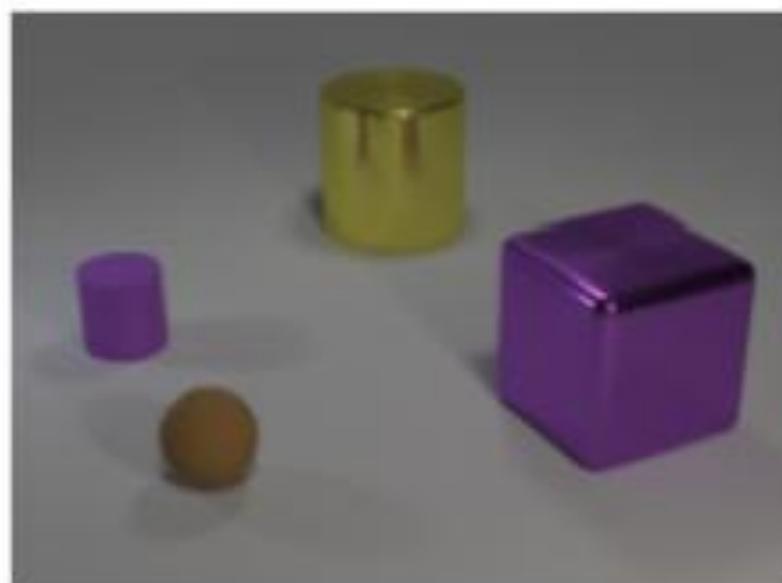
Somewhat related to the image

Unrelated to the image

## Понимание изображений

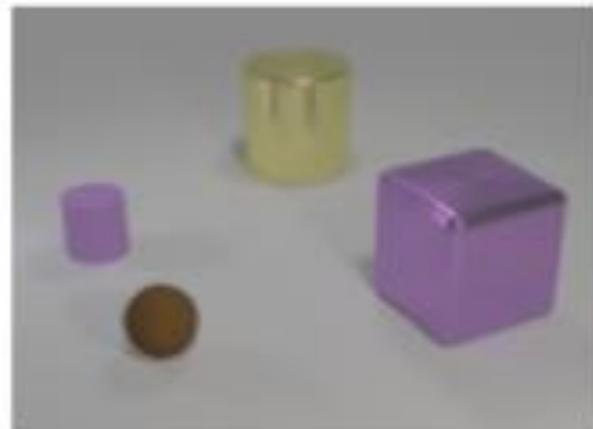
### Visual reasoning

Original Image:



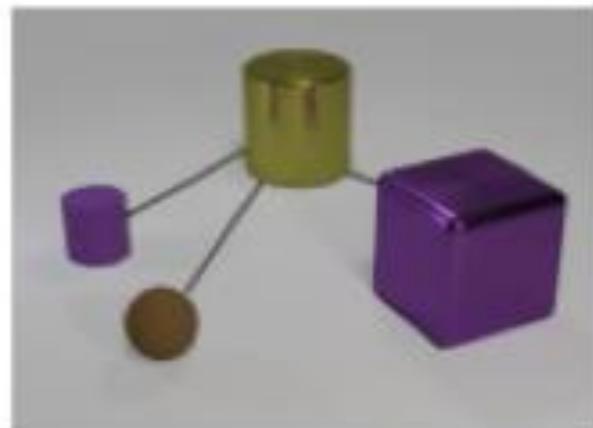
**Non-relational question:**

What is the size of the brown sphere?



**Relational question:**

Are there any rubber things that have the same size as the yellow metallic cylinder?



<https://arxiv.org/pdf/1706.01427.pdf>

## Понимание изображений Visual QA Challenge

**Соревнование <https://visualqa.org/roe.html>**  
**вопросы двух форматов: OpenEnded / MultipleChoice**

Who is wearing glasses?

man



woman

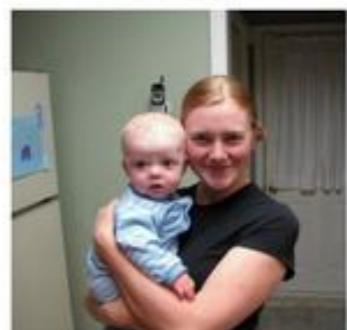


Where is the child sitting?

fridge



arms



Is the umbrella upside down?

yes



no



How many children are in the bed?

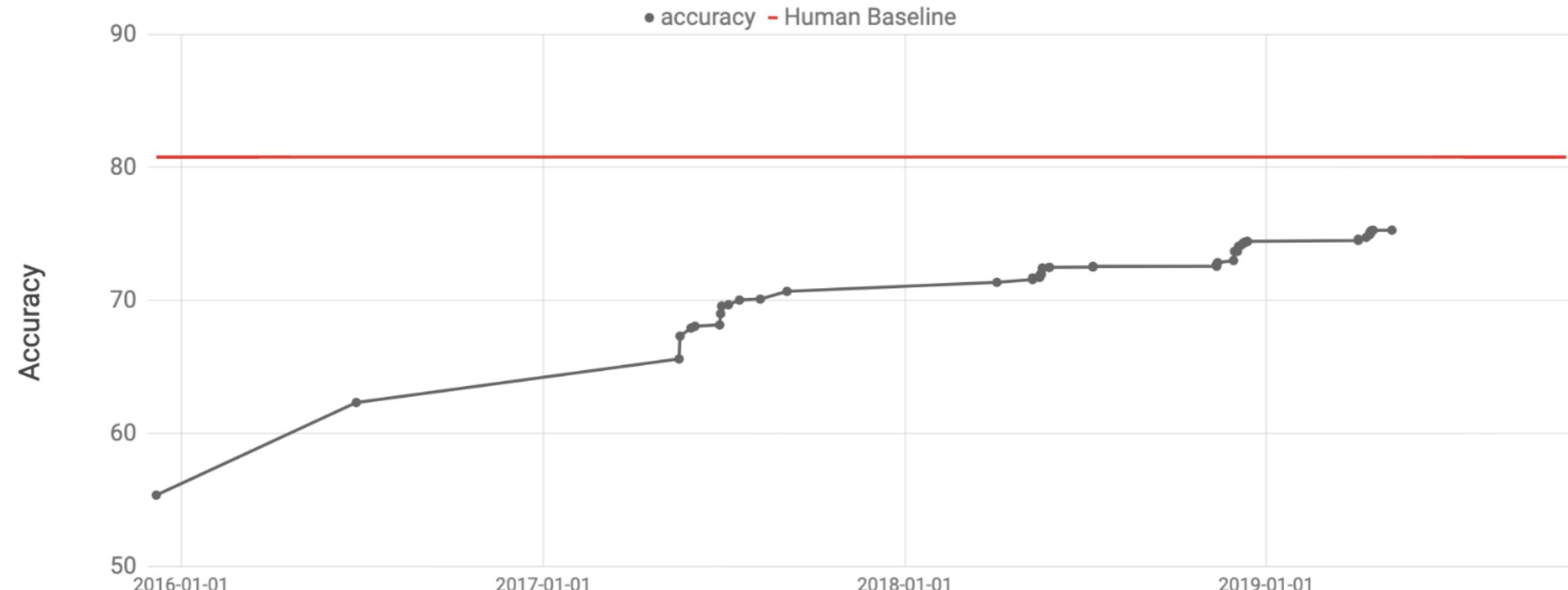
2



1



## Понимание изображений Visual QA Challenge



Note: Human performance is measured by having humans answer questions for images and evaluating their answers using the same metrics as we use to evaluate machines that answer the same questions. Inter-human disagreement, paraphrased answers, spelling errors, etc, contribute to human performance being (quite a bit lower) than 100%.

[Artificial Intelligence Index Report 2019]

## Понимание изображений

### объяснение изображений (Visual Explanations)

*This is a Black-Capped Vireo because...*



Description: this bird has a white belly and breast black and white wings with a white wingbar.

Explanation-Dis: this is a bird with a white belly yellow wing and a **black head**.

*This is a Crested Auklet because...*



Description: this bird is black and white in color with a orange beak and black eye rings.

Explanation-Dis.: this is a black bird with a **white eye** and an orange beak.

*This is a White Pelican because...*



Description: this bird is white and black in color with a long curved beak and white eye rings.

Explanation: this is a large white bird with a **long neck** and a **large orange beak**.

*This is a Geococcyx because...*



Description: this bird has a long black bill a white throat and a brown crown.

Explanation-Dis.: this is a black and white spotted bird with a **long tail feather** and a pointed beak.

<https://arxiv.org/pdf/1603.08507.pdf>

## Текст в изображение (text2image)

Text descriptions  
(content)

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.

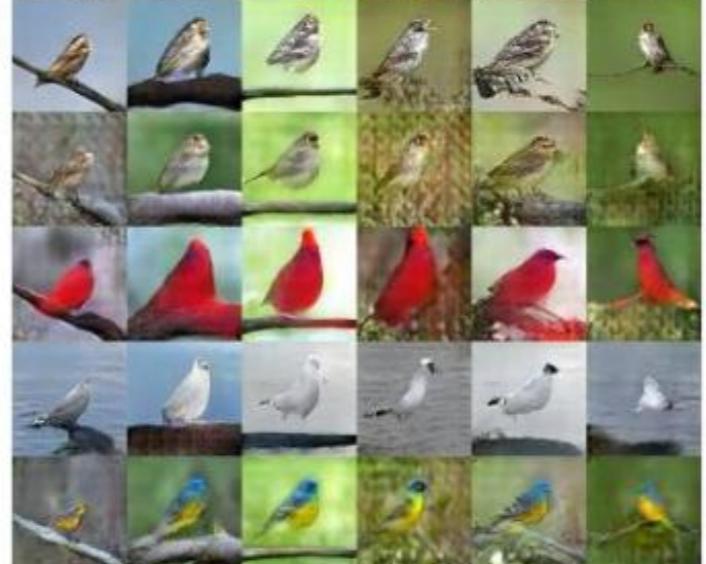
A small bird with **white base** and **black stripes** throughout its belly, head, and feathers.

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>

# Текст в изображение (text2image)

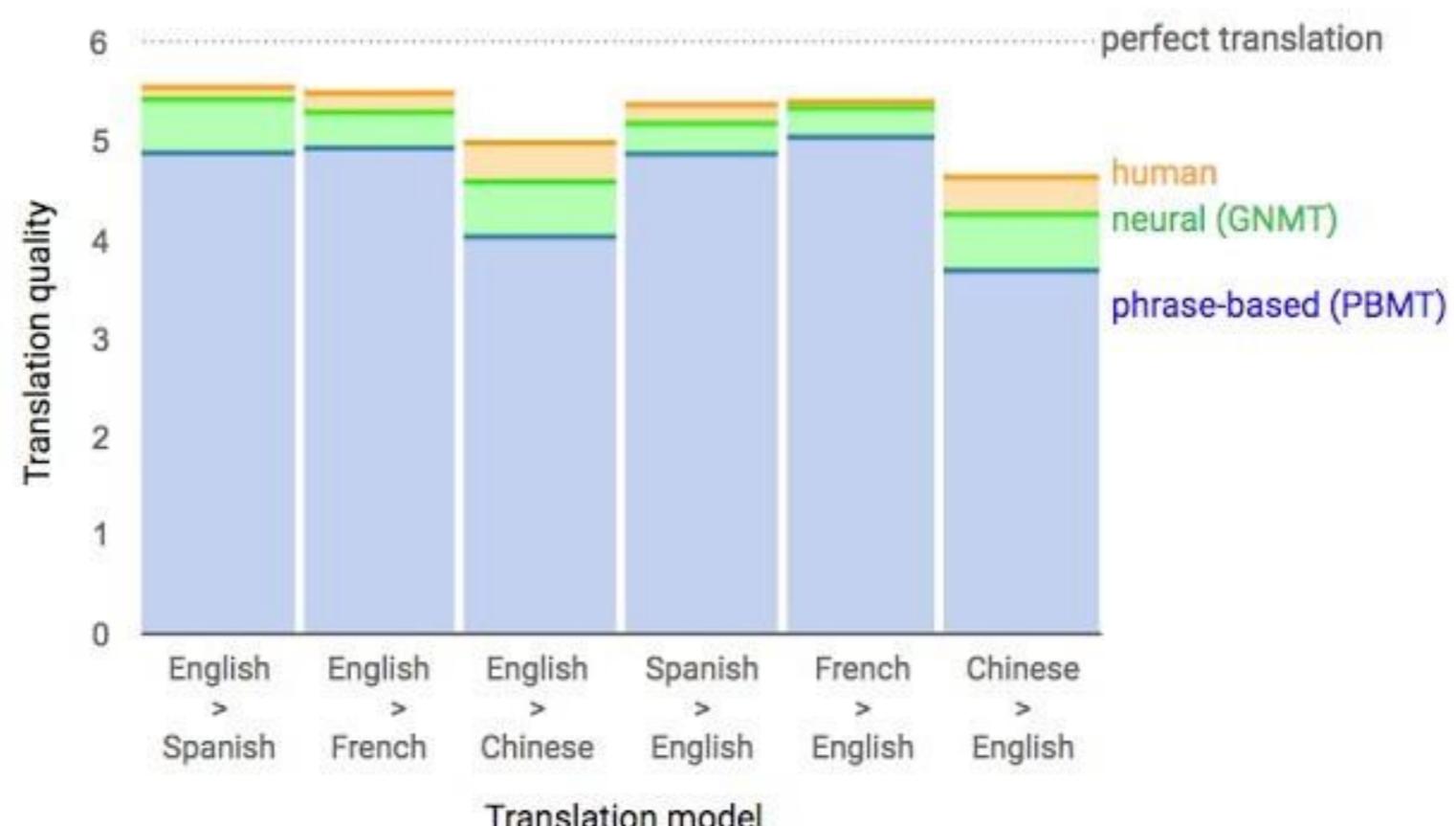
**Генерация по запросу (здесь: чайник в стиле кубика Рубика)**



<https://openai.com/blog/dall-e/>

## Обработка текста (Natural Language Processing)

- Понимание смысла (Language Understanding)
  - Автоматический перевод

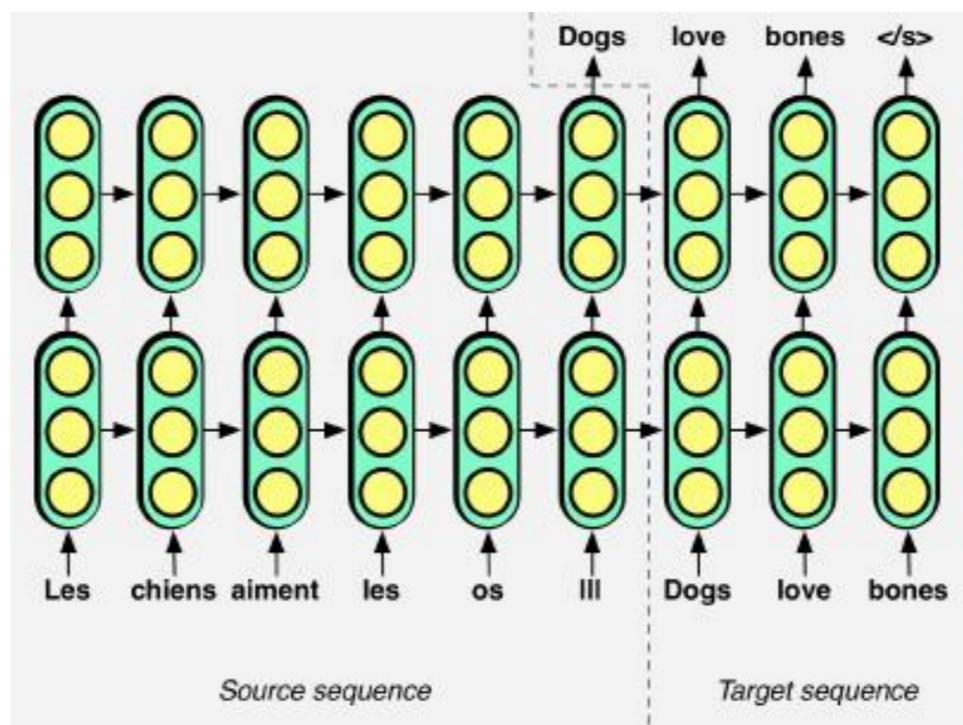


Google's Neural Machine Translation System

<https://arxiv.org/abs/1609.08144>

## Обработка текста (Natural Language Processing)

- Диалоговые системы



Иван вчера утром поехал на работу, но встретил Петю и весь день провёл в кино, а сегодня его уволили.

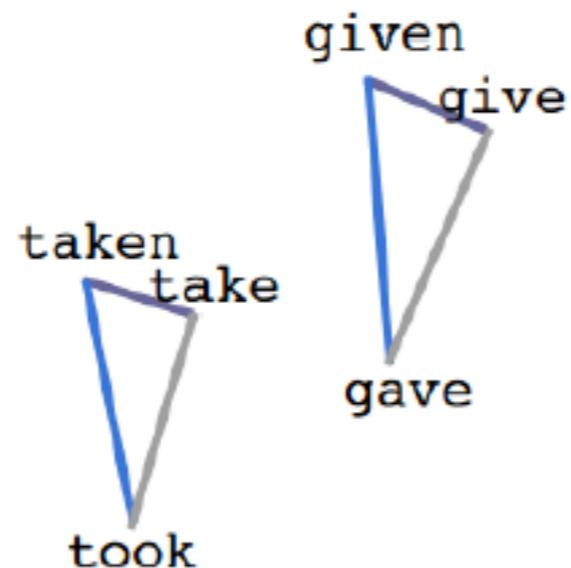
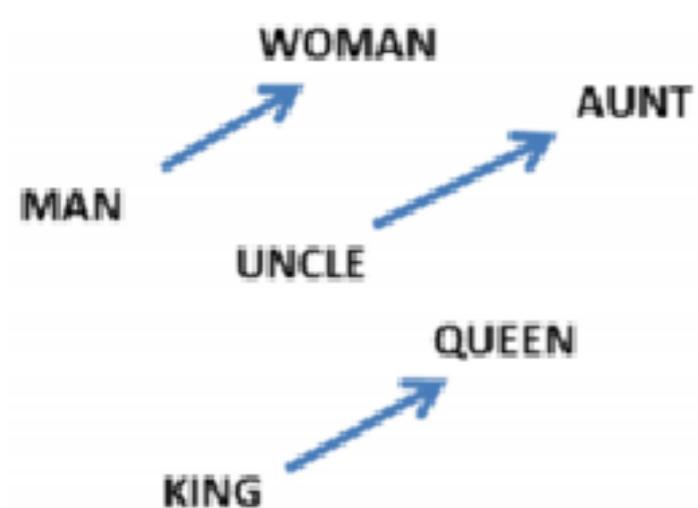
Что делал вчера Иван? **Был в кино.**

Кого Петя вчера встретил? **Ивана.**

Сегодня \_\_\_\_\_ того, кто вчера не работал, а был в кино. **уволили**

## Обработка текста (Natural Language Processing)

### Представления слов, предложений, текстов...



Woman – Man  $\approx$  Aunt - Uncle  
King - Male + Female  $\approx$  Queen  
Human - Animal  $\approx$  Ethics

## Языковые модели / генерация текста



**Сегодня мы начали курс по глубокому обучению. В нём будет много разного – в том числе нефизических приёмов, которые очень быстро приведут тебя в состояние Сверхчеловека. Позже ты узнаешь, как им пользоваться.**

**Что нас не убивает - делает сильнее. А что убивает - само становится сильным, ибо никогда не теряется. Вчерашнее убийство человека есть слабость - сегодняшнее убьет того, кто уничтожит того, кто его уничтожит. Кто эти мгновения?**

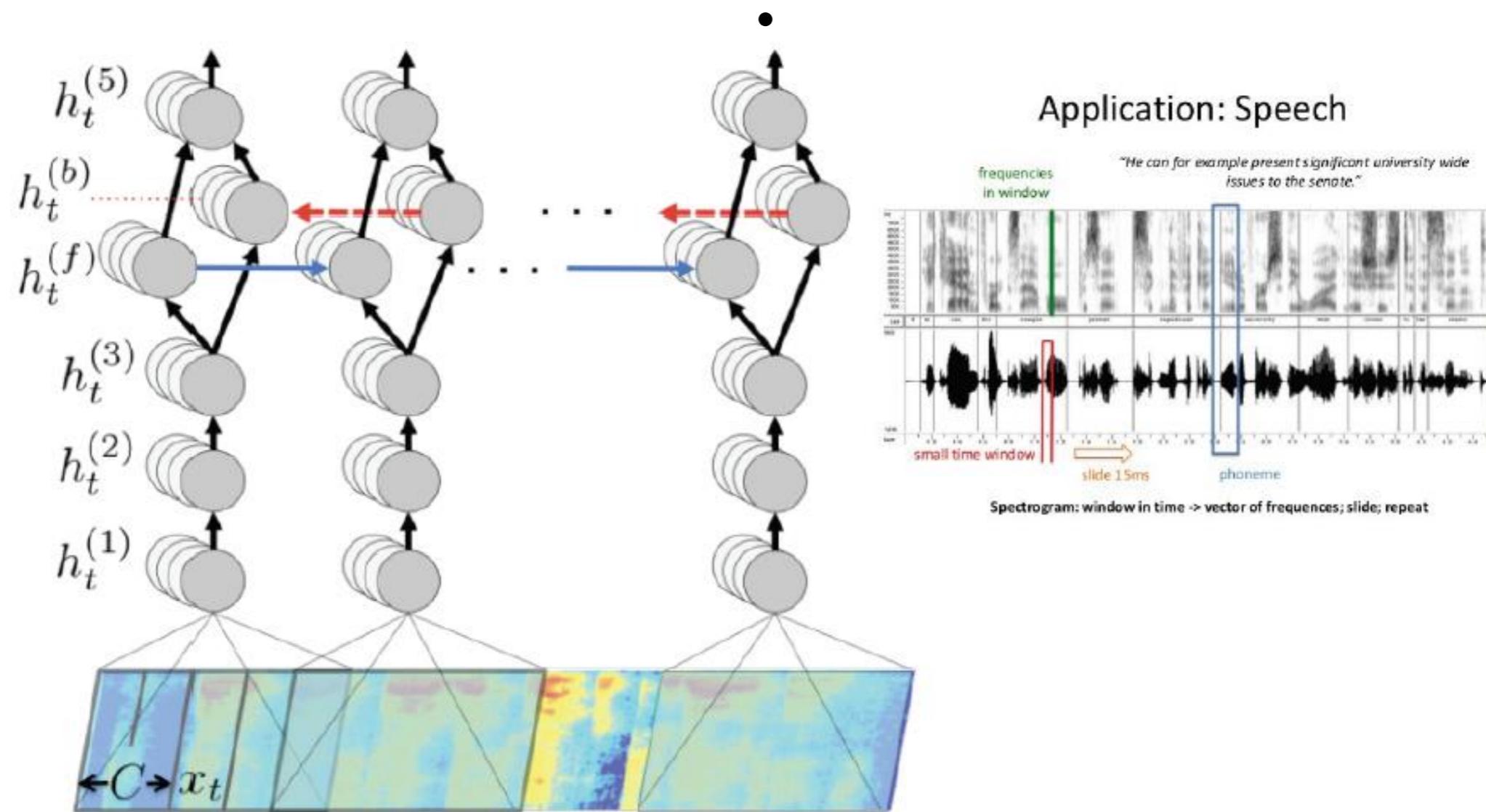
**Мумми-троль посмотрел на Шрека и предложил ему виски. Я не пью этот сорт грустно сказал Шрек и обнял Пяточка. Доехав до Большого Парка, он поставил «бьюик» возле скверика, выключил мотор и задремал, свалив затылок на плюшевую подушечку.**

**Умножив всё равенство на ноль, можно заметить, что всё сократилось, даже интеграл, который не брался аналитически. Отсюда можно сделать вывод: Вселенная должна была быть меньше в тысячу раз, чем мы предполагаем. Теперь попробуем выяснить, какова же она на самом деле.**

<https://porfirevich.ru/>

## Обработка аудио / звуков / речи

- Распознавание речи (Speech Recognition) – уже было



## Обработка видео / аудио

### Чтение по губам (Lip reading)



<https://arxiv.org/pdf/1611.05358v1.pdf>

### Синхронизация движения губ



[http://grail.cs.washington.edu/projects/AudioToObama/siggraph17\\_obama.pdf](http://grail.cs.washington.edu/projects/AudioToObama/siggraph17_obama.pdf)

DataSet: [http://www.robots.ox.ac.uk/~vgg/data/lip\\_reading/lrs2.html](http://www.robots.ox.ac.uk/~vgg/data/lip_reading/lrs2.html)

## Обработка видео / аудио: Deepfake



## Обработка видео / аудио: Deepfake



<https://style.rbc.ru/impressions/5d6909ea9a7947497ec5cca2>

## Обработка видео / аудио: Генерация речи

### Tacotron 2 (Google)

- произношение
  - громкость
  - скорость
  - интонация

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

### Melotron

(based on Tacotron 2 GST,  
NVIDIA)

- трансфер ритма
- голос из аудио

<https://arxiv.org/abs/1910.11997>

### Fastpitch (NVIDIA)

- параллельный
- обусловленный
- трансформер

<https://fastpitch.github.io/>

### VITS (Variational Inference with adversarial learning for end-to-end Text-to-Speech)

- параллельный

- end2end

- stochastic duration predictor

<https://arxiv.org/pdf/2106.06103.pdf>

## Автоматизация колл-центров / разговорный ИИ



Главная Вакансии Блог

+7 (495) 134-81-11

# Послушайте, как наш робот совершает звонки

▶ Распределение входящих  
звонков дилерского  
центра  
00:38

▶ Обзвон ледяной базы  
лидов для квалификации  
00:43

▶ Продление платного  
размещения вакансии  
01:17

**предложение услуг / обратная связь / решение проблем / спам-фильтр**

## Синтез музыки

### Warner Music signs first ever record deal with an algorithm

Mood music app Endel, which creates bespoke soundscapes for users, is expected to produce 20 albums this year



см. также <https://aiva.ai/>

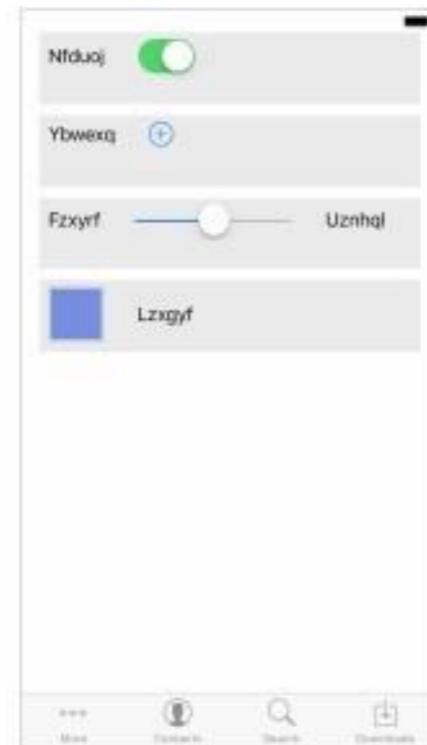
## Автоматизация программирования: конвертирование кода

Python input	Unsupervised C++ translation
<pre> def SumOfKsubArray(arr, n, k):     Sum = 0     S = deque()     G = deque()     for i in range(k):         while (len(S) &gt; 0 and arr[S[-1]] &gt;= arr[i]):             S.pop()         while (len(G) &gt; 0 and arr[G[-1]] &lt;= arr[i]):             G.pop()         G.append(i)         S.append(i)     for i in range(k, n):         Sum += arr[S[0]] + arr[G[0]]         while (len(S) &gt; 0 and S[0] &lt;= i - k):             S.popleft()         while (len(G) &gt; 0 and G[0] &lt;= i - k):             G.popleft()         while (len(S) &gt; 0 and arr[S[-1]] &gt;= arr[i]):             S.pop()         while (len(G) &gt; 0 and arr[G[-1]] &lt;= arr[i]):             G.pop()         G.append(i)         S.append(i)         Sum += arr[S[0]] + arr[G[0]]     return Sum </pre>	<pre> int SumOfKsubArray(int arr[], int n, int k){     int Sum = 0;     deque &lt;int&gt; S;     deque &lt;int&gt; G;     for(int i = 0; i &lt; k; i ++){         while((int) S.size() &gt; 0 &amp;&amp; arr[S.back()] &gt;= arr[i])             S.pop_back();         while((int) G.size() &gt; 0 &amp;&amp; arr[G.back()] &lt;= arr[i])             G.pop_back();         G.push_back(i);         S.push_back(i);     }     for(int i = k; i &lt; n; i ++){         Sum += arr[S.front()] + arr[G.front()];         while((int) S.size() &gt; 0 &amp;&amp; S.front() &lt;= i - k)             S.pop_front();         while((int) G.size() &gt; 0 &amp;&amp; G.front() &lt;= i - k)             G.pop_front();         while((int) S.size() &gt; 0 &amp;&amp; arr[S.back()] &gt;= arr[i])             S.pop_back();         while((int) G.size() &gt; 0 &amp;&amp; arr[G.back()] &lt;= arr[i])             G.pop_back();         G.push_back(i);         S.push_back(i);     }     Sum += arr[S.front()] + arr[G.front()];     return Sum; } </pre>

Figure 2: Example of unsupervised Python to C++ translation. TransCoder successfully translates the Python input function `SumOfKsubArray` into C++. TransCoder infers the types of the arguments, of the variables, and the return type of the function. The model maps the Python `deque()` container, to the C++ implementation `deque<>`, and uses the associated `front`, `back`, `pop_back` and `push_back` methods to retrieve and insert elements into the `deque`, instead of the Python square brackets `[]`, `pop` and `append` methods. Moreover, it converts the Python `for` loop and `range` function properly.

**unsupervised-обучение на GitHub-е** <https://arxiv.org/pdf/2006.03511.pdf>

## Автоматизация программирования: Pix2Code



(a) iOS GUI screenshot

```
stack {  
    row {  
        label, switch  
    }  
    row {  
        label, btn-add  
    }  
    row {  
        label, slider, label  
    }  
    row {  
        img, label  
    }  
    footer {  
        btn-more, btn-contact, btn-search, btn-download  
    }  
}
```

(b) Code describing the GUI written in our DSL

Figure 2: An example of a native iOS GUI written in our DSL.

**есть также исправление ошибок в коде**  
<https://arxiv.org/abs/2005.10636>

## ИИ в играх: Atari, шахматы, го, дота...



**2013, 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!**

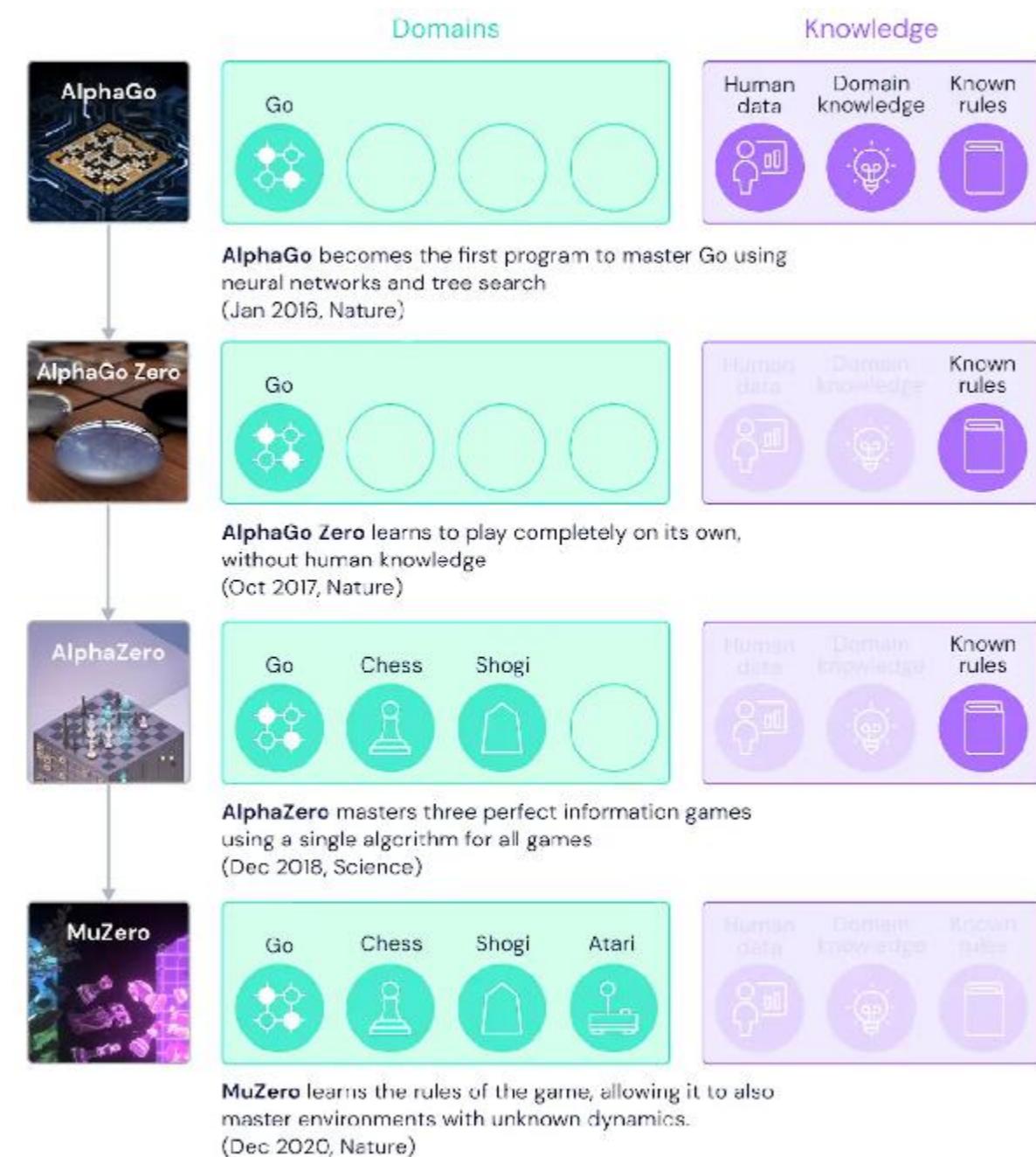
**9-15.03.2016**

**AlphaGo (Google DeepMind) — Ли Седоль**

**4 – 1**

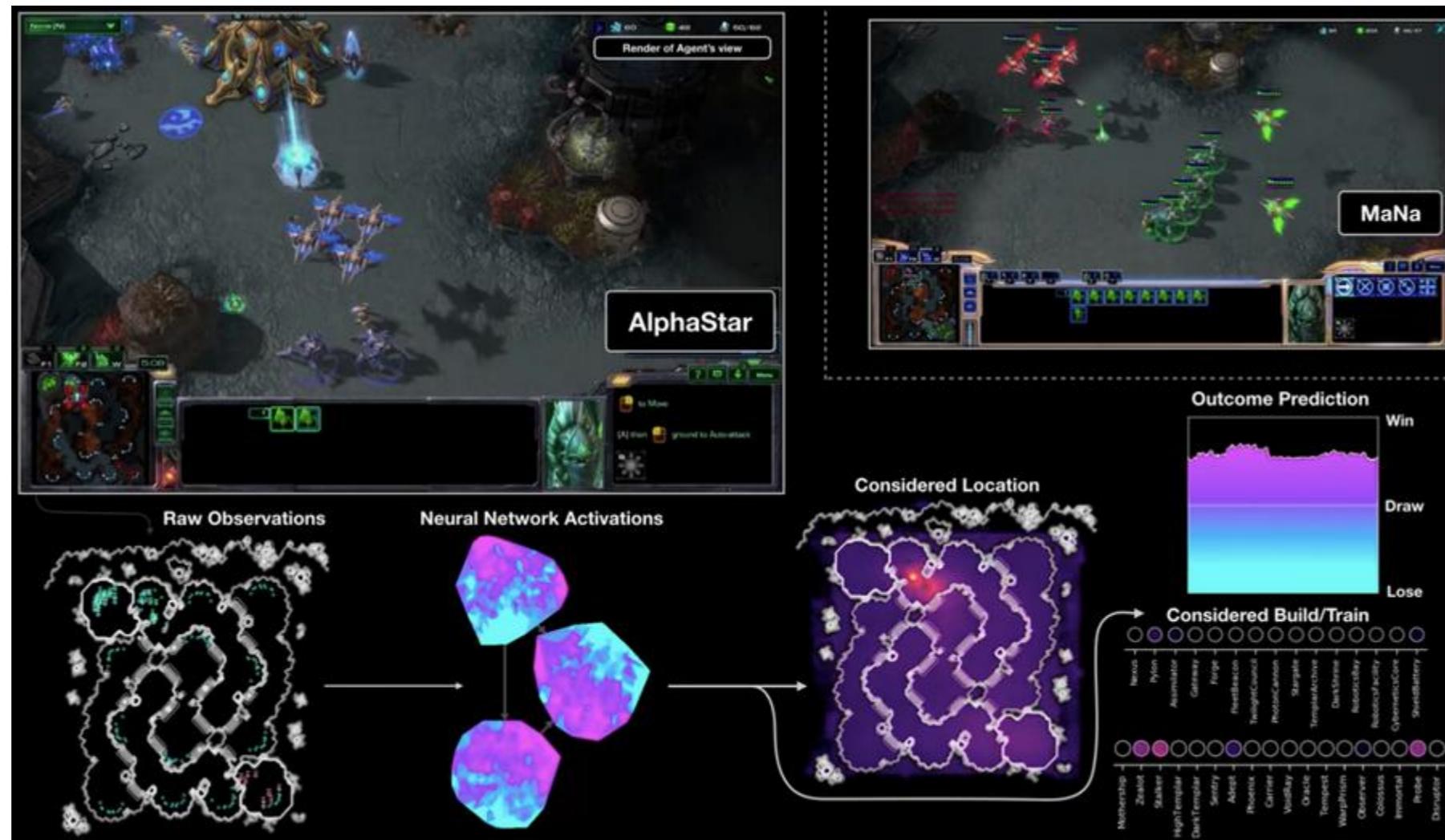
**Предсказание действия и оценки позиции**

**AlphaGo – 1920 CPUs, 280 GPUs, электричество = 3000\$**



<https://deepmind.com/blog/article/muzero-mastering-go-chess-shogi-and-atari-without-rules>

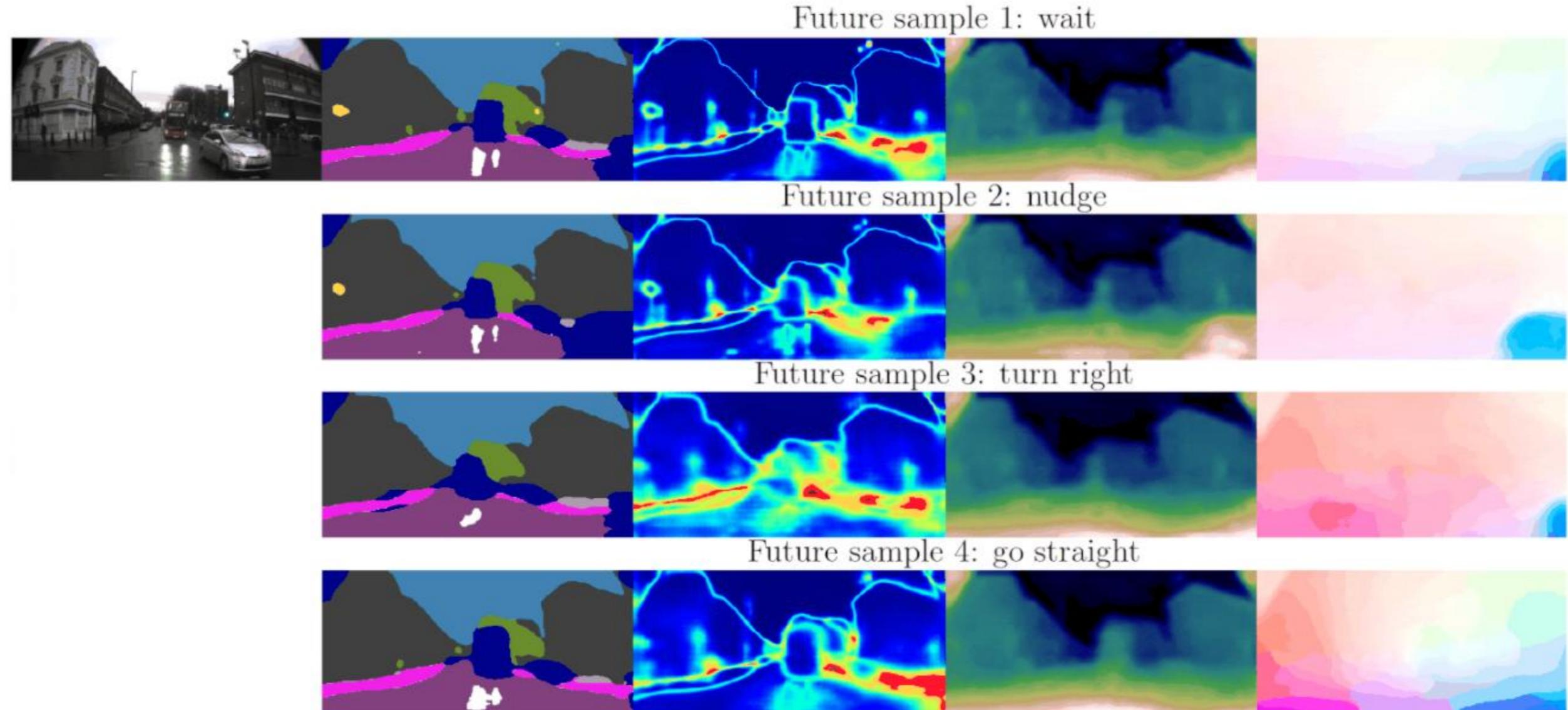
## AlphaStar от DeepMind для StarCraft II



19/12/12 – AlphaStar – Grzegorz Komincz (MaNa) из команды Liquid – 5:0

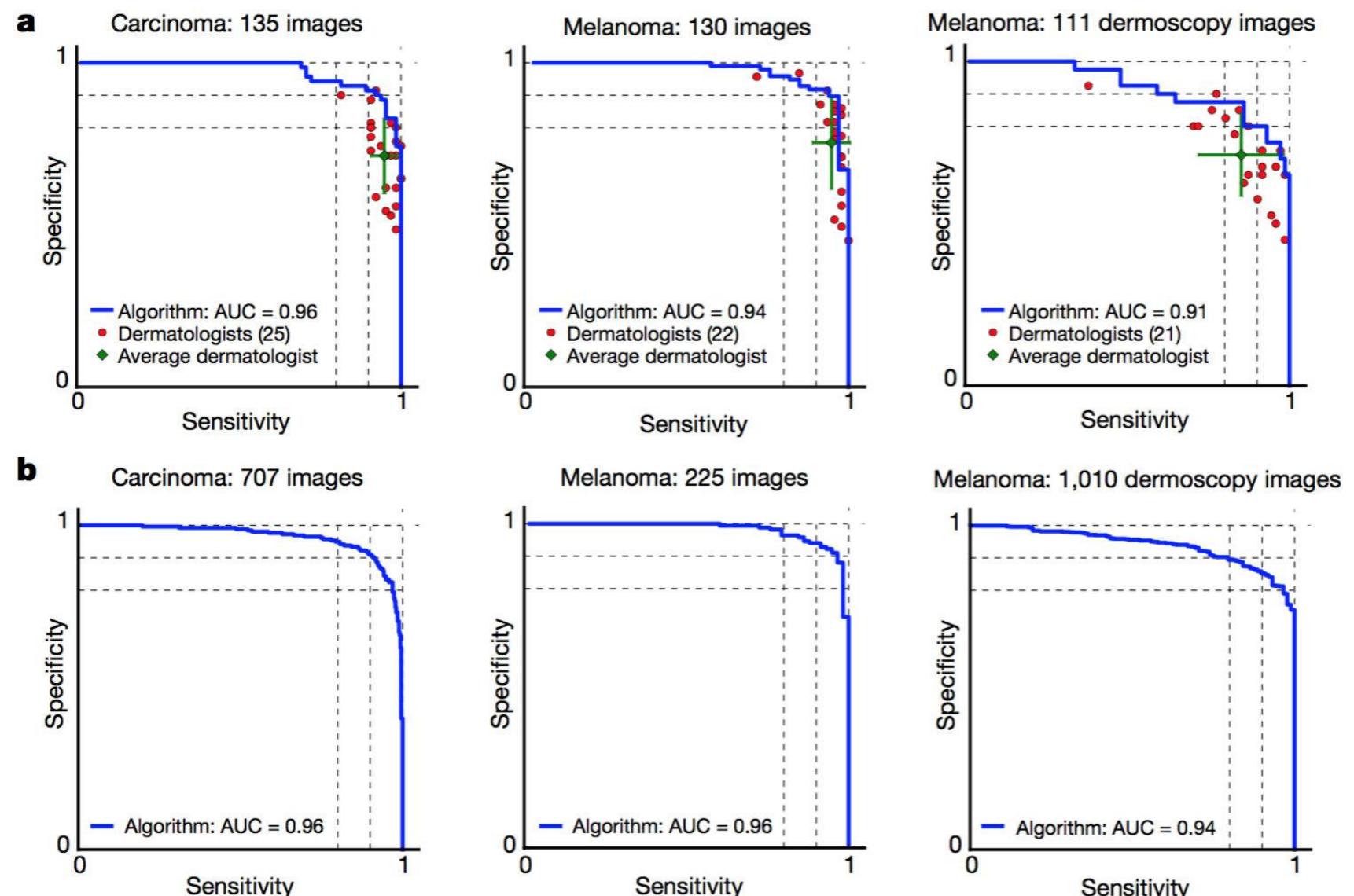
<https://habr.com/ru/post/437486/>

## RL: Предсказание ситуации на дороге



<https://arxiv.org/pdf/1911.09070.pdf>

## Медицина (Drug Discovery and Medical): диагностика рака кожи



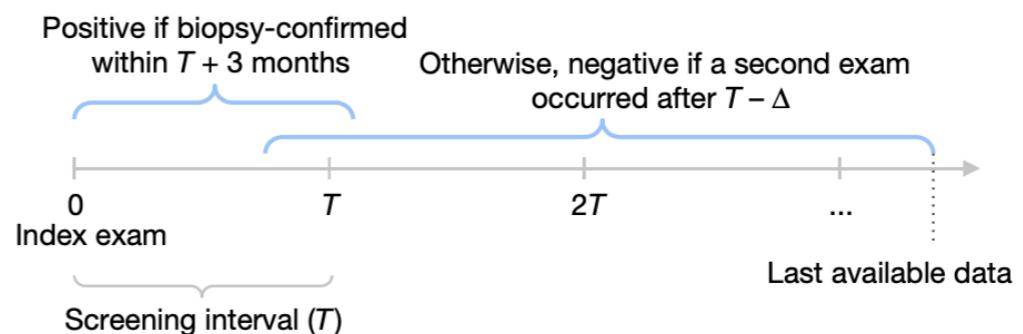
<https://cs.stanford.edu/people/esteva/nature/>

# Медицина (Drug Discovery and Medical): скрининг мамограмм

## Test datasets

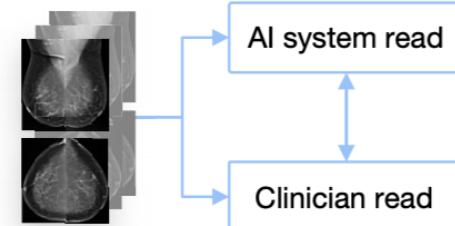
Number of women	25,856	3,097
Interpretation	Double reading	Single reading
Screening interval	3 years	1 or 2 years
Cancer follow-up	39 months	27 months
Number of cancers	414 (1.6%)	686 (22.2%)

## Ground-truth determination



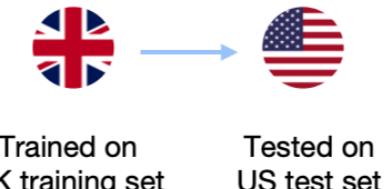
## Evaluation

Comparison with retrospective clinical performance

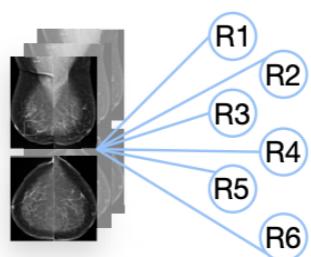


UK and  
US test sets

Generalization  
across datasets



Independently conducted  
reader study

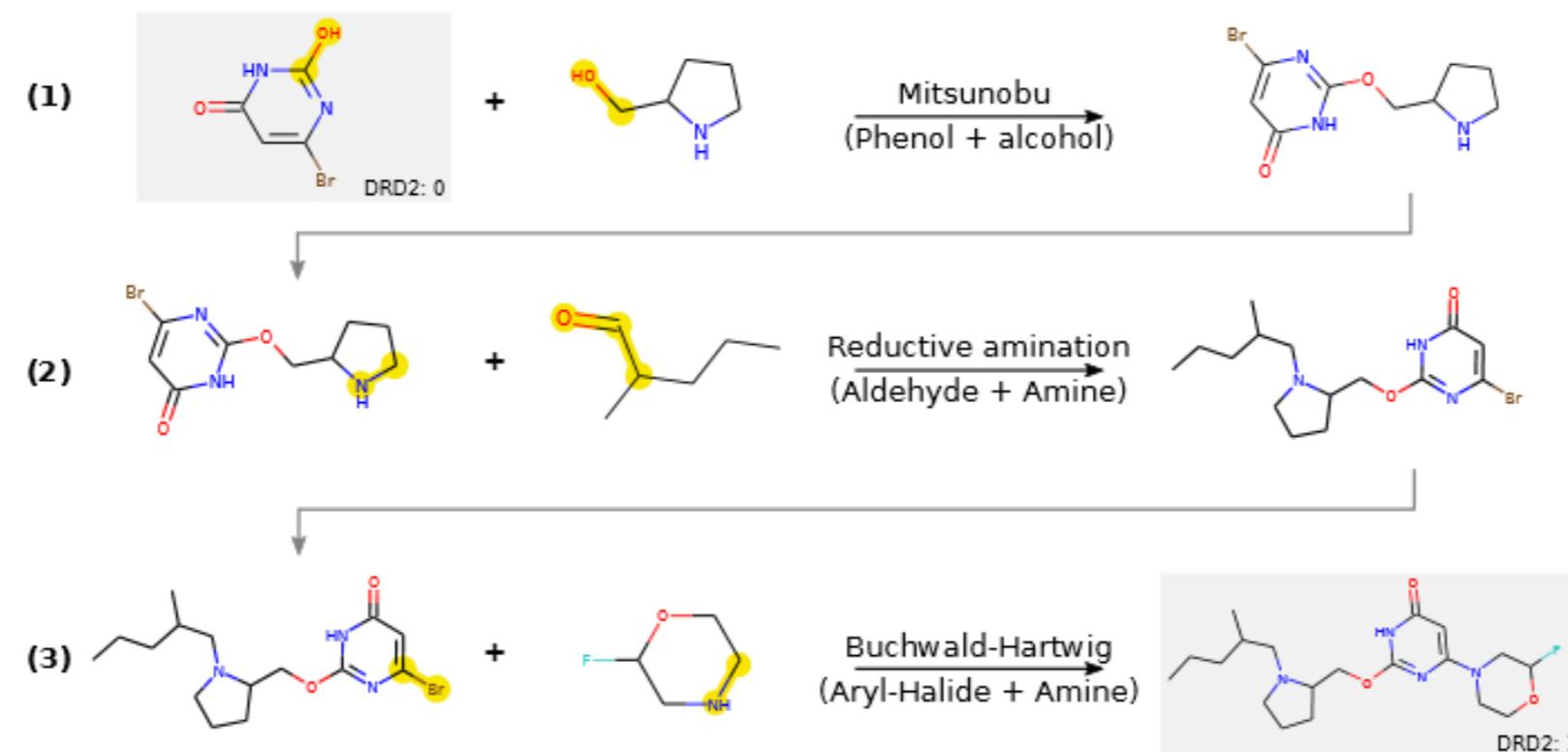


6 radiologists read 500 cases  
from US test set

<https://www.nature.com/articles/s41586-019-1799-6>

## Медицина (Drug Discovery and Medical)

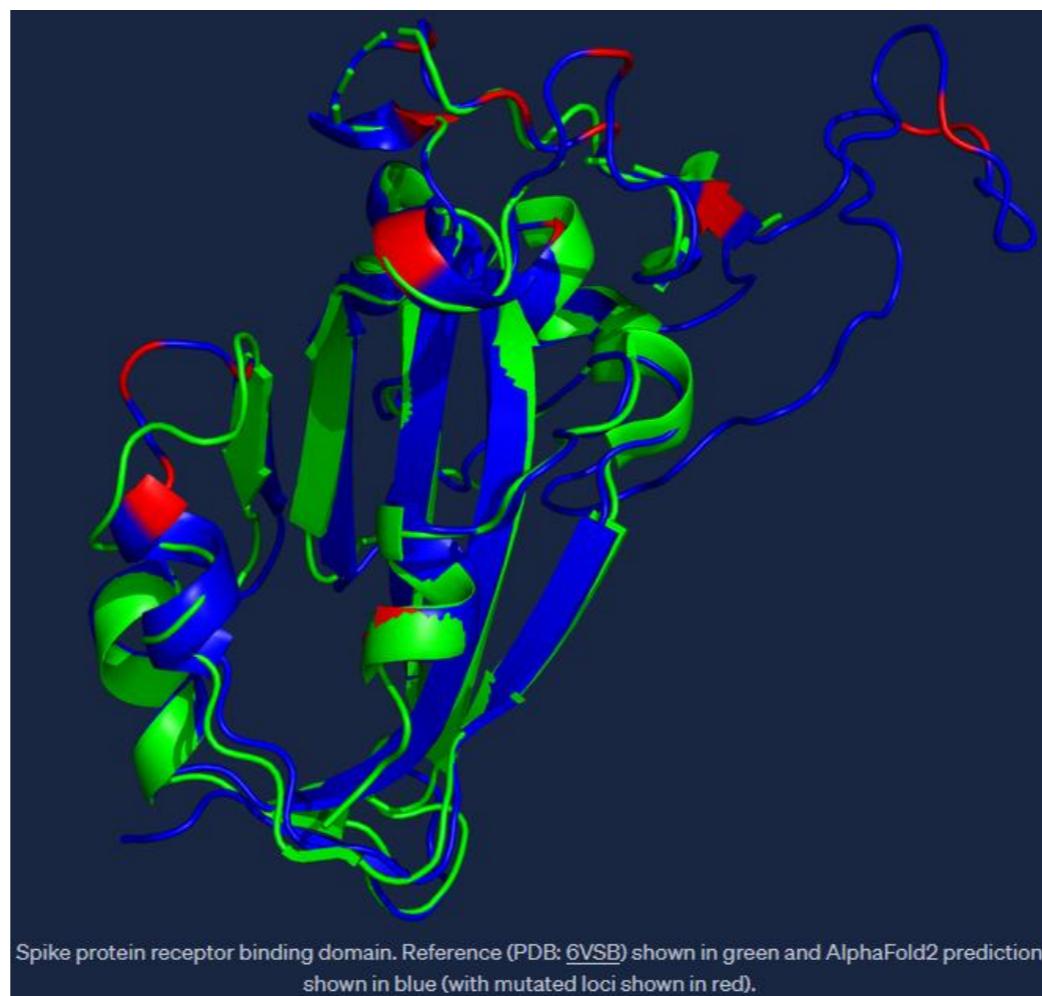
### Предсказание свойств химических соединений с помощью обучения с подкреплением



Julien Horwood «Molecular Design in Synthetically Accessible Chemical Space via Deep Reinforcement Learning» // <https://arxiv.org/pdf/2004.14308.pdf>

## Медицина и химия

### Предсказание структуры белка и его свойств



<https://colbyford.medium.com/protein-structure-prediction-of-b-1-1-529-sars-cov-2-spike-variant-with-alphafold2-39c5bf9cf9ed>

## Наука: автоматические доказательства

### PROBLEM 1

*Adapted from AMC12 2000 Problem 5*

Prove that if  $|x - 2| = p$ , where  $x < 2$ , then  $x - p = 2 - 2p$ .

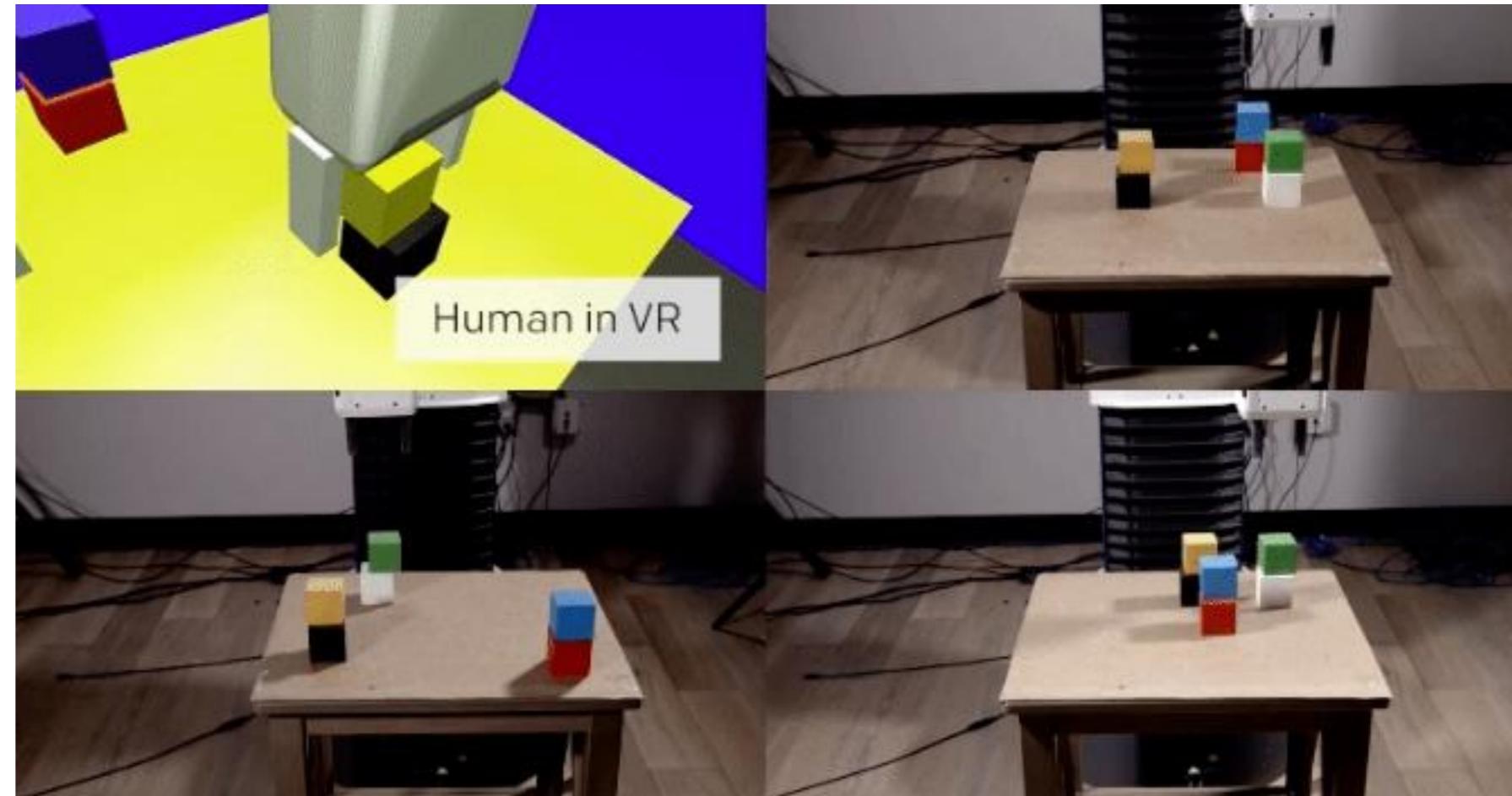
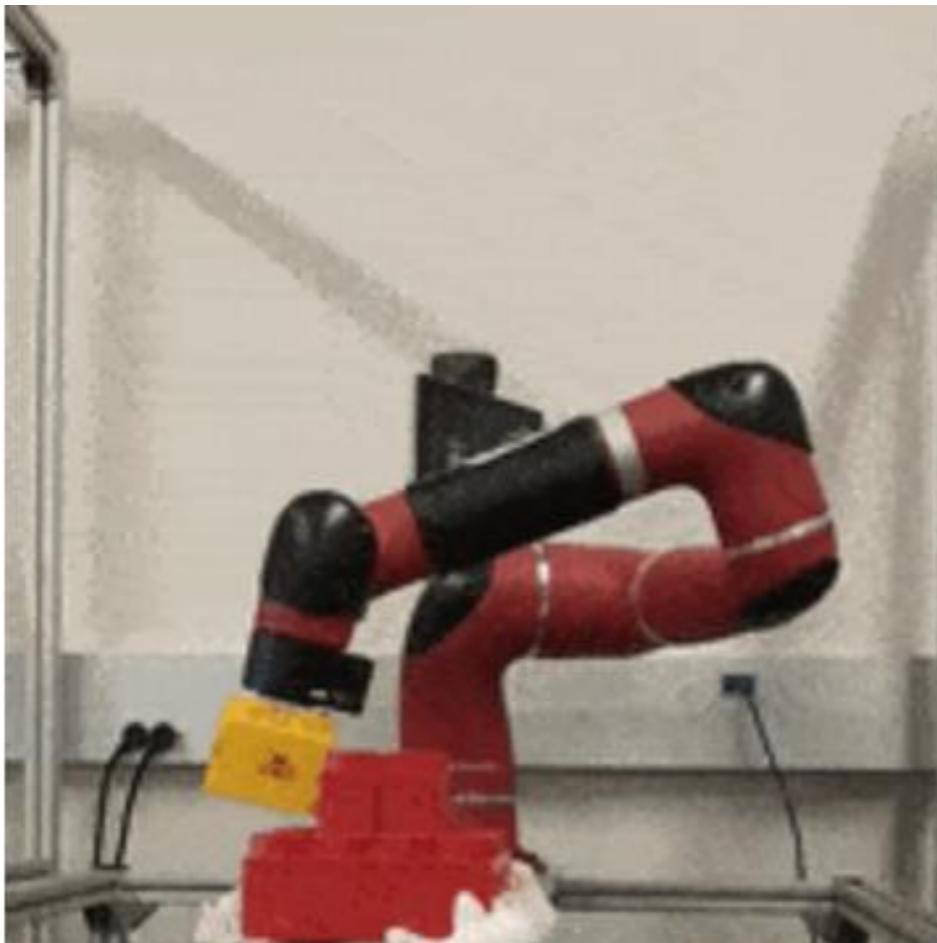
◊ FORMAL

INFORMAL

Since  $x < 2$ ,  $|x - 2| = -(x - 2)$ . Using  $p = |x - 2|$  we have  $x = 2 - p$  and finally  $x - p = 2 - 2p$ .

<https://openai.com/blog/formal-math/>

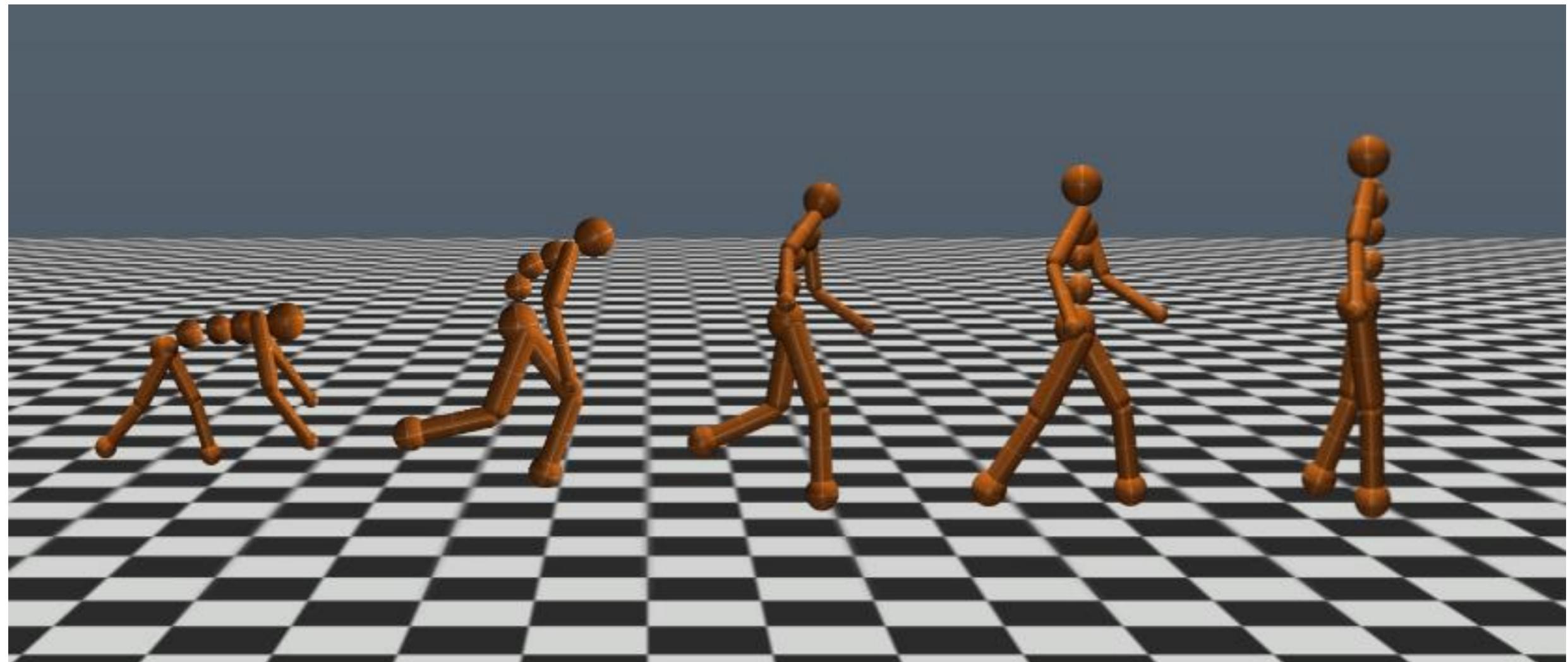
## Автомобили-беспилотники, роботы-манипуляторы



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

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

## Нейроэволюционные алгоритмы



<https://eng.uber.com/deep-neuroevolution/>

## Подбор гардероба

Most compatible



Least compatible



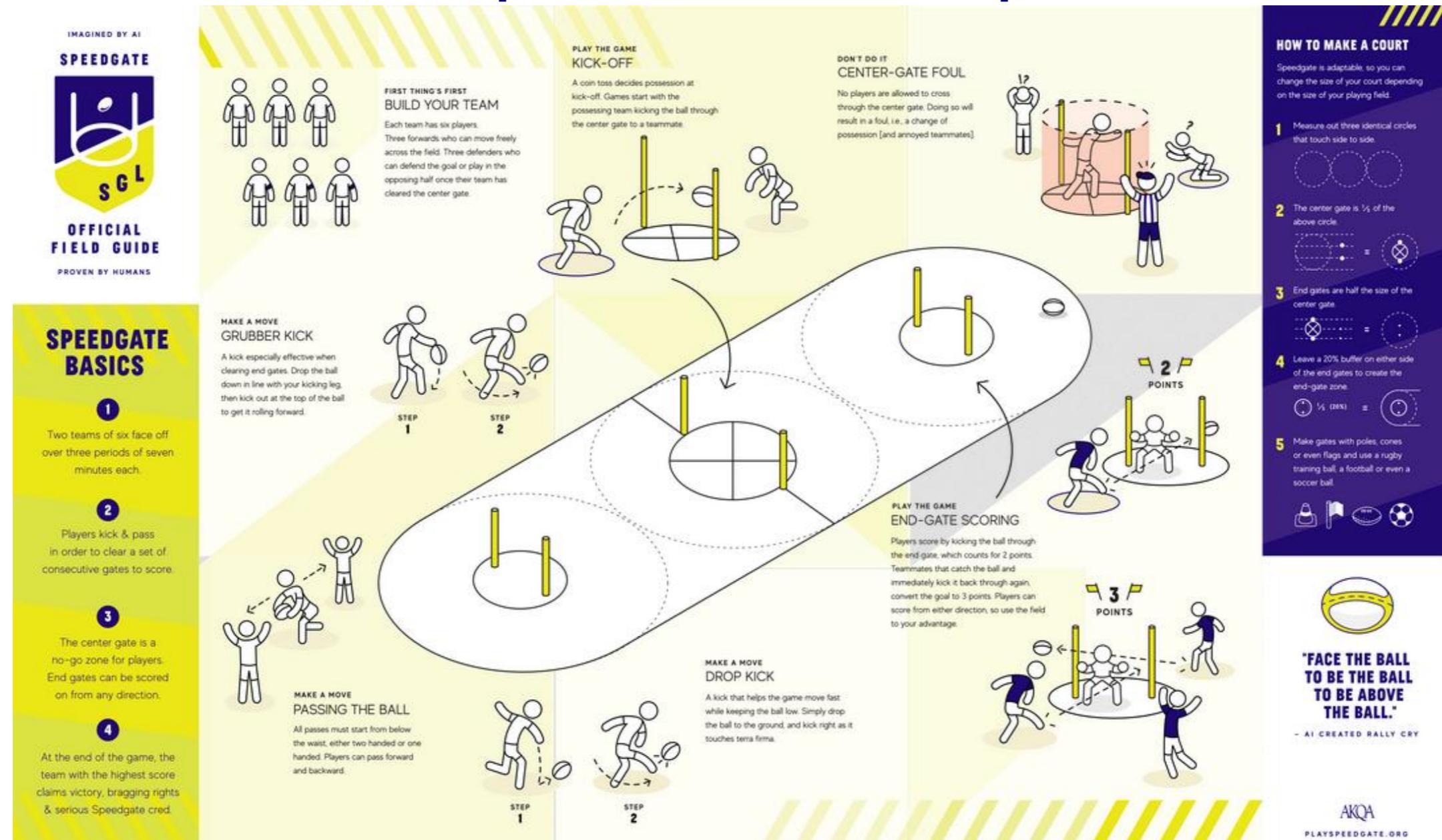
<https://arxiv.org/pdf/1712.02662.pdf>

## Подбор гардероба



<https://colab.research.google.com/github/tg-bomze/collection-of-notebooks/blob/master/HomeStylist.ipynb#scrollTo=AsrgeAyzzQ-v>

# Изобретение нового вида спорта



<https://techcrunch.com/2019/04/11/akqa-says-it-used-ai-to-invent-a-new-sport-called-speedgate/>

## Моделирование спортивных поединков

### Моделирование игроков, траектории мяча, синтез поединков между любыми соперниками

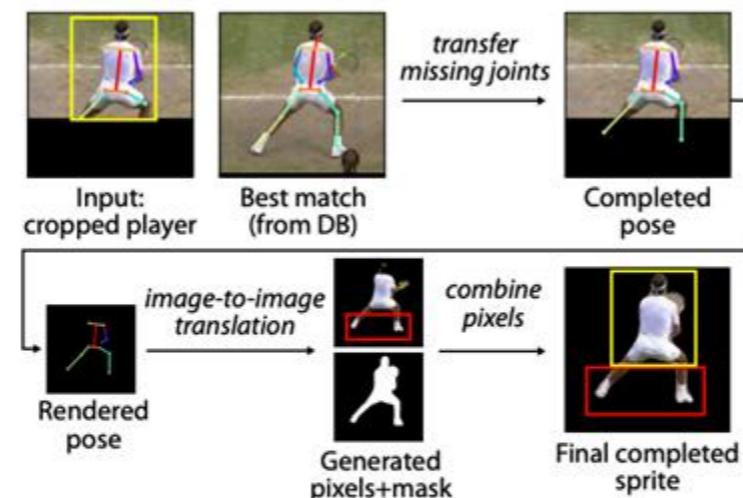
Target: images from short time window (same lighting conditions)



Source: images under different lighting conditions



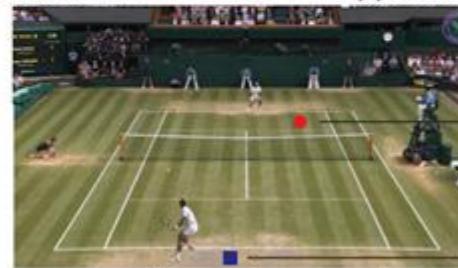
Transformed source images (lighting matches target)



(a) Williams vs. Federer



(b) Federer vs. Federer



(c) Interactive control

User specified shot placement

User specified recovery position

<https://cs.stanford.edu/~haotianz/research/vid2player/>

## AutoML: пусть DL делает всё!

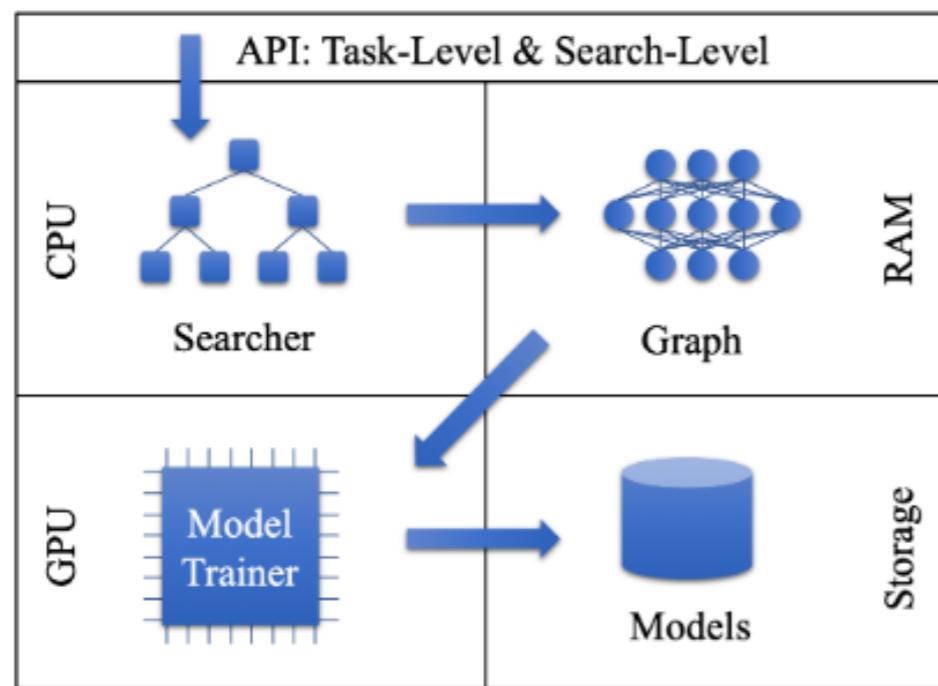


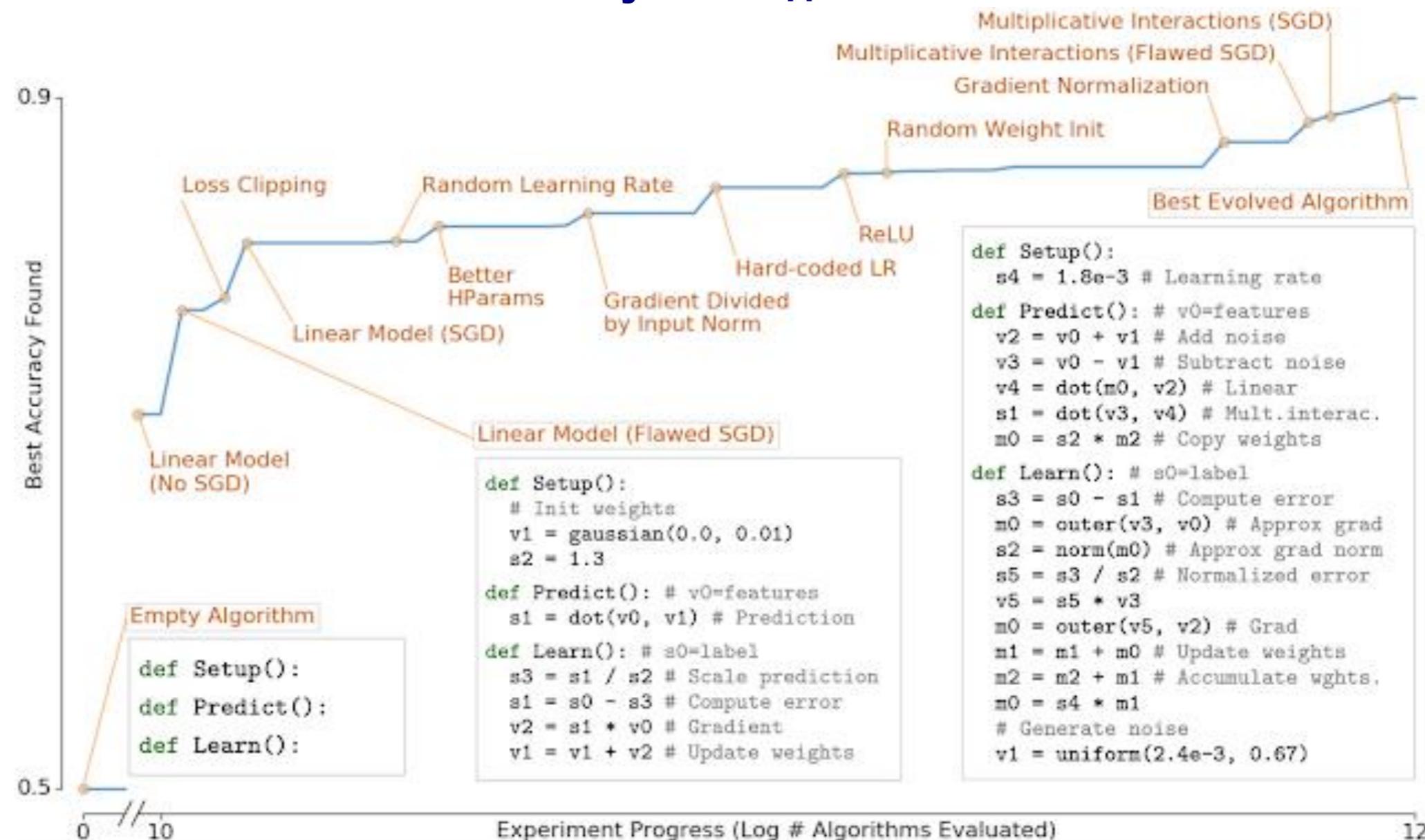
Figure 2: Auto-Keras System Overview. (1) User calls the API. (2) The Searcher generates neural architectures on CPU. (3) Graph builds real neural networks with parameters on RAM from the neural architectures. (4) The neural network is copied to GPU for training. (5) Trained neural networks are saved on storage devices.

Table 1: Classification Error Rate

Methods	MNIST	CIFAR10	FASHION
RANDOM	1.79%	16.86%	11.36%
GRID	1.68%	17.17%	10.28%
SPMT	1.36%	14.68%	9.62%
SMAC	1.43%	15.04%	10.87%
SEAS	1.07%	12.43%	8.05%
NASBOT	NA	12.30%	NA
BFS	1.56%	13.84%	9.13%
BO	1.83%	12.90%	7.99%
AK	0.55%	11.44%	7.42%
AK-DP	0.60%	3.60%	6.72%

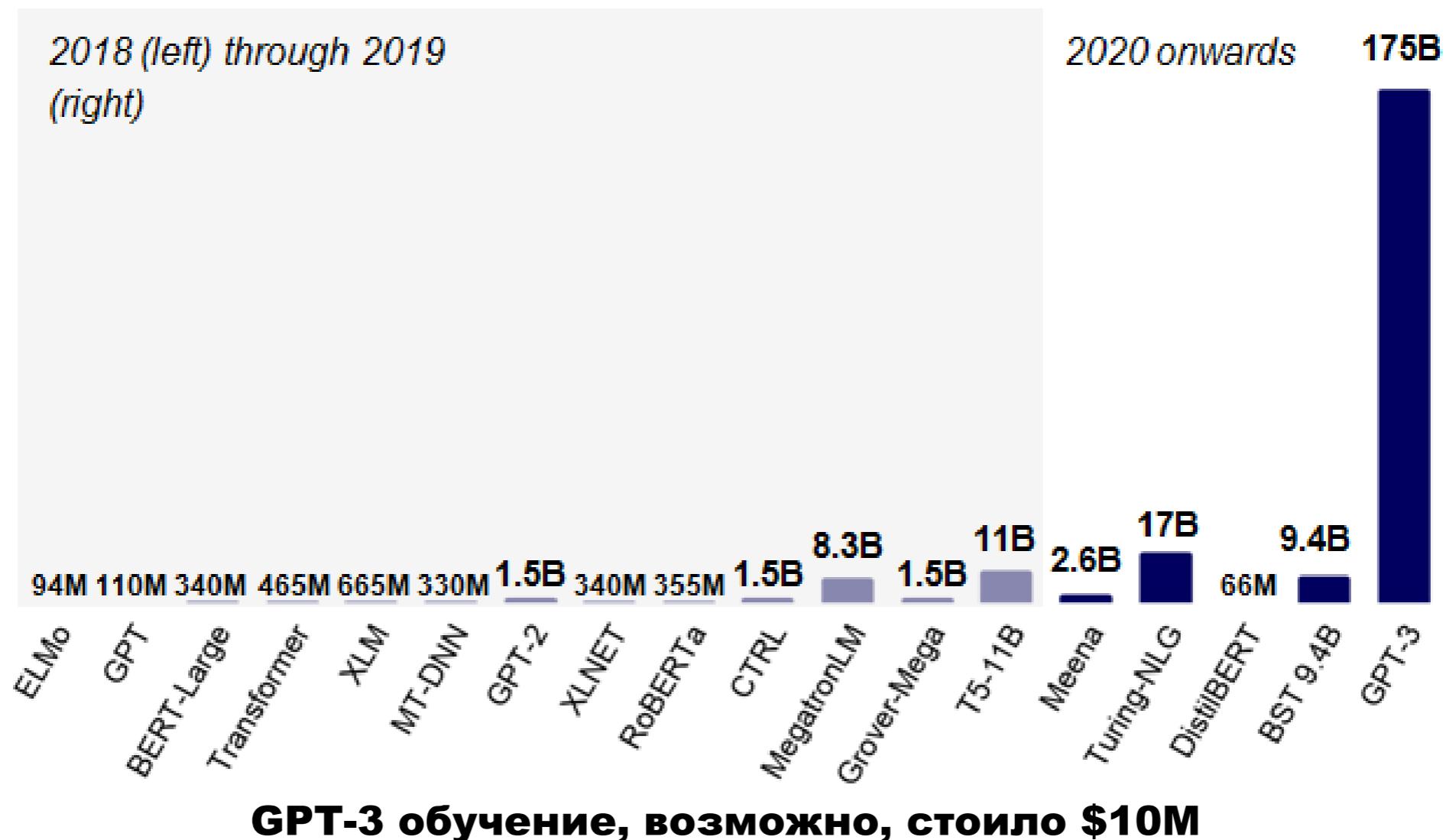
«AutoKeras» <https://arxiv.org/pdf/1806.10282.pdf>

## AutoML: пусть DL делает всё!



<https://ai.googleblog.com/2020/07/automl-zero-evolving-code-that-learns.html>

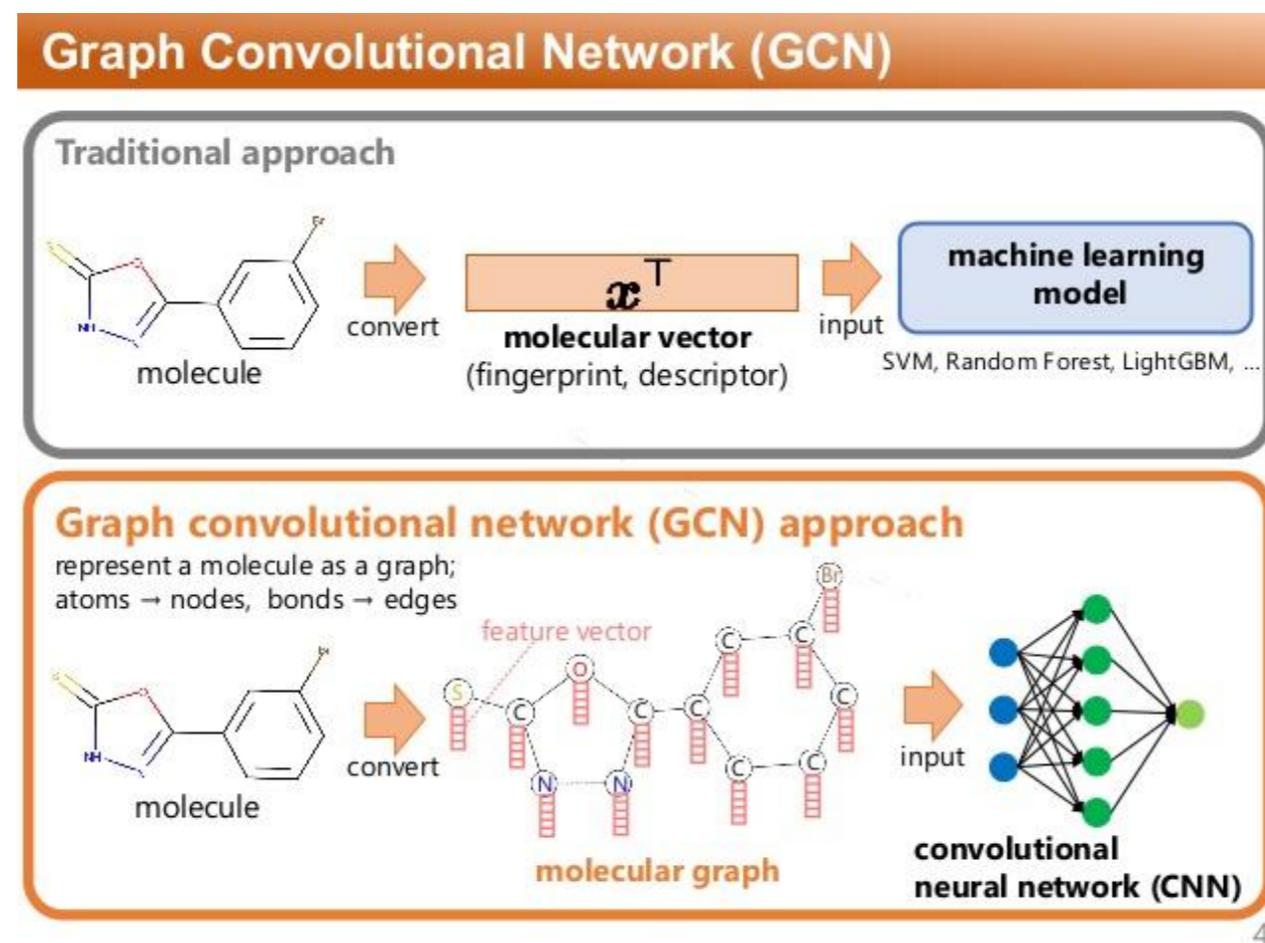
## Тренды: большие модели



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

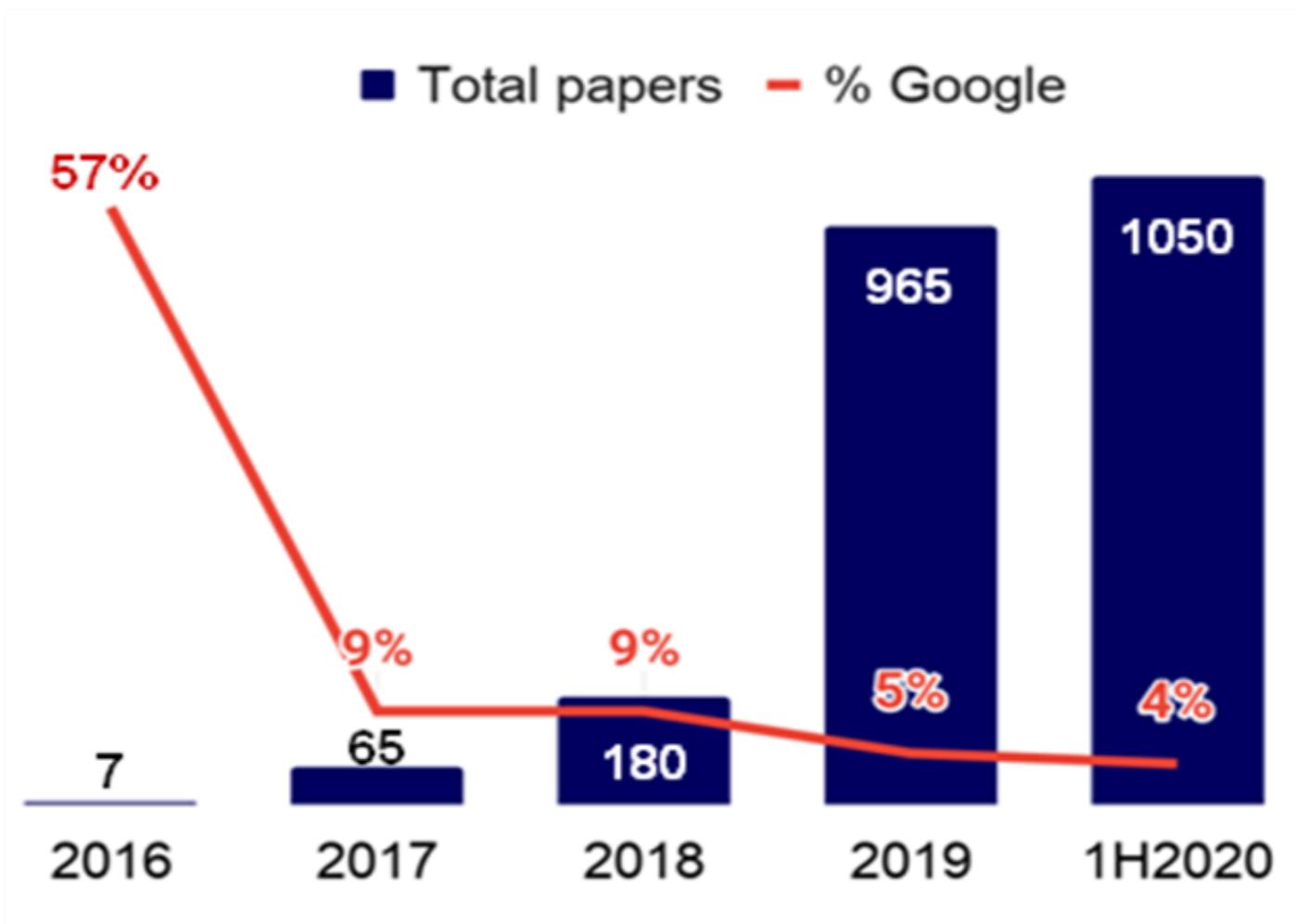
## Тренды: новые архитектуры и парадигмы

### Появление новых архитектур сетей, отход от традиционных признаковых описаний



<https://www.slideshare.net/tonets/pdpta19ohue>

## Тренды: федеративное обучение (Federated Learning)

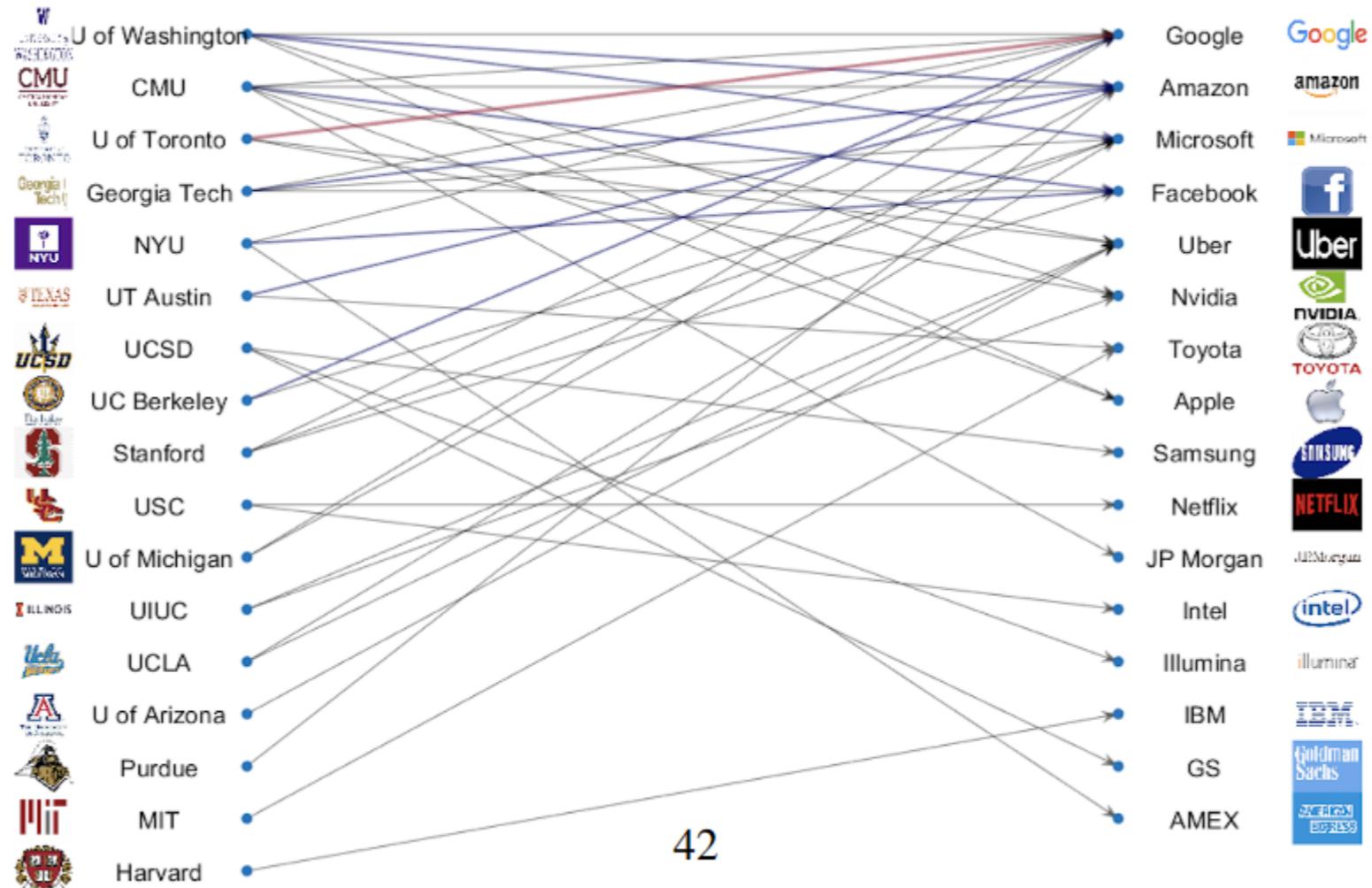


**Децентрализованное обучение на разных устройствах**

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

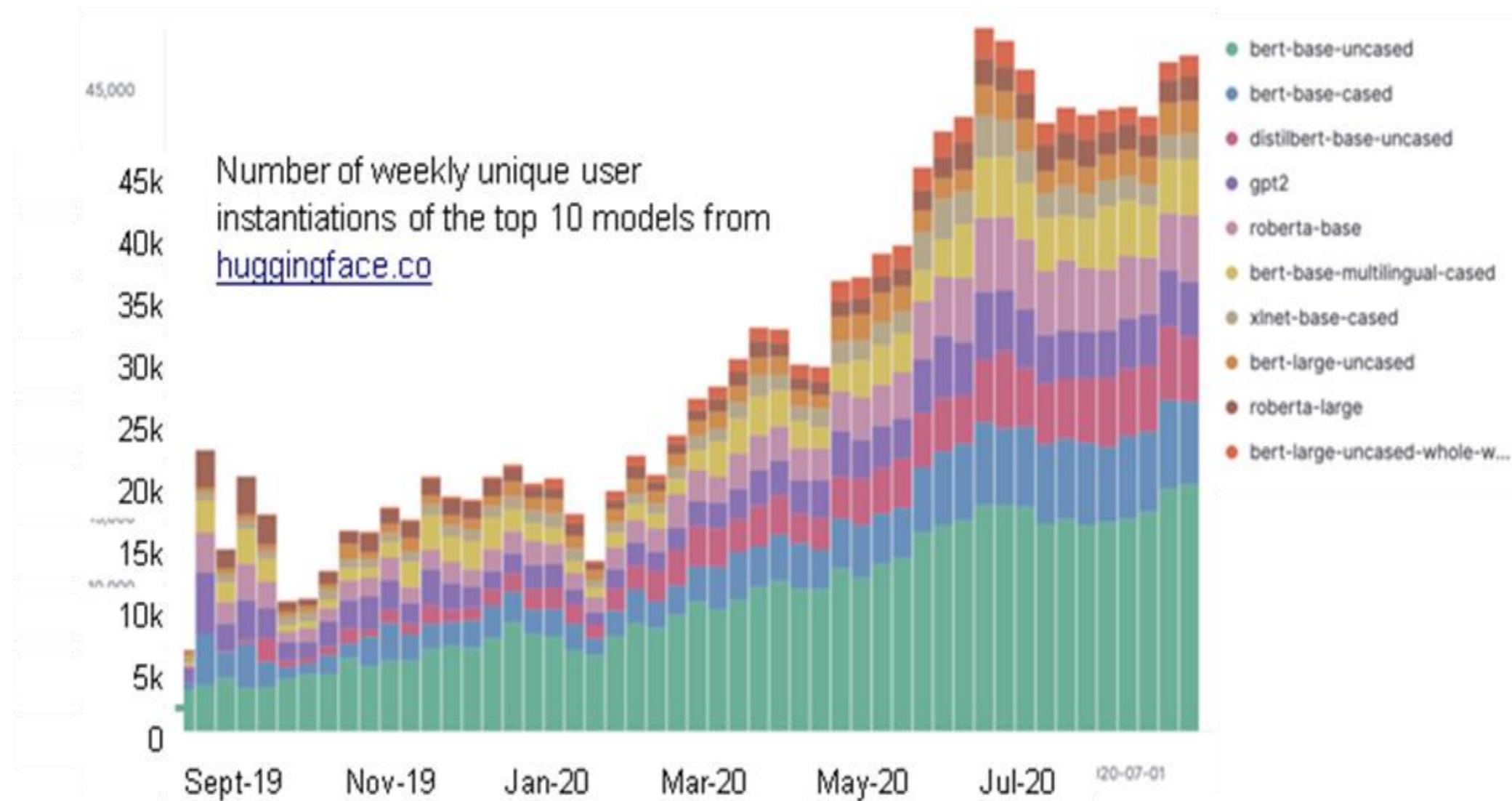
## Трены: академическое сообщество «переезжает» в компании

**Кто откуда хантит профессоров:**



[http://gofman.info/AI/AI\\_GofmanJin.pdf](http://gofman.info/AI/AI_GofmanJin.pdf)

## Тренды: хорошие открытые библиотеки



**Больше 1000 компаний используют библиотеку от Hugging Face**

## Причины прогресса

- Интенсивные исследования в ML (+ обмен идеями)
- Огромные наборы данных (+ интернет, дататификация, миниатюризация)
- Hardware (CPUs, GPUs + всё новое специальное)
- Новые приёмы в обучении (Dropout, BN и т.п.)
- Архитектурные: возможность инвариантности к некоторым преобразованиям (например масштабу) + принципиально новые идеи (ex: Transformer)
- Интерес бизнеса

Графический процессор	Год выпуска	Техпроцесс	Гигафлопсов, для 32-разрядных вычислений	Гигафлопсов, для 64-разрядных вычислений
Geforce RTX 2080 Ti (TU102)	2018	12 нм	13450	420,2
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	$2 \times 2550 = 5100$	$2 \times 637 = 1274$
Radeon HD 5970 (2xRV870)	2009	40 нм	$2 \times 2320 = 4640$	$2 \times 464 = 928$
Radeon R9 Fury X (Fuji)	2015	28 нм	8602	538

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

## Проблемы и минусы нейросетей

- **Переобучение, большое число весов**
  - Разделение параметров
  - Использование структуры данных (архитектуры)
  - Регуляризация
- **Нестабильное обучение, долгое обучение, плохие решения**
  - новые методы оптимизации
  - эвристические приёмы
  - архитектуры
  - GPU, кластеры GPU
- **Необходимость больших данных**
  - Каудсорс-разметка
  - Предобученные модели
  - Самообучение
  - One-shot-подходы
  - Эффективные представления (embeddings – стилей, интонации и т.п.)

## Проблемы и минусы нейросетей

- Не учитывают «естественный контекст», слабая интерпретация решений
  - Пока не ИИ...
  - Не учат инварианты (альтернативы – капсулевые сети и т.п.)
  - Теория интерпретаций
- Уязвимости (возможность атак)
  - Подходы к защите сетей
- Ресурсоёмкость
  - Сети облегчённой архитектуры (для мобильных приложений)
  - Обучение «Учитель-ученик»
- Проблемы с выбором архитектуры
  - Нет теории...
  - NAS (автоматический подбор)

сеть	ошибка	время
<b>ResNet18</b>	<b>0.108</b>	<b>2.5 дня</b>
<b>ResNet50</b>	<b>0.07</b>	<b>5 дней</b>
<b>ResNet101</b>	<b>0.062</b>	<b>1 неделя</b>

## Датасеты

Year	Breakthroughs in AI	Datasets (First Available)	Algorithms (First Proposed)
1994	Human-level spontaneous speech recognition	Spoken Wall Street Journal articles and other texts (1991)	Hidden Markov Model (1984)
1997	IBM Deep Blue defeated Garry Kasparov	700,000 Grandmaster chess games, aka "The Extended Book" (1991)	Negascout planning algorithm (1983)
2005	Google's Arabic- and Chinese-to-English translation	1.8 trillion tokens from Google Web and News pages (collected in 2005)	Statistical machine translation algorithm (1988)
2011	IBM Watson became the world Jeopardy! champion	8.6 million documents from Wikipedia, Wiktionary, Wikiquote, and Project Gutenberg (updated in 2010)	Mixture-of-Experts algorithm (1991)
2014	Google's GoogLeNet object classification at near-human performance	ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010)	Convolution neural network algorithm (1989)
2015	Google's Deepmind achieved human parity in playing 29 Atari games by learning general control from video	Arcade Learning Environment dataset of over 50 Atari games (2013)	Q-learning algorithm (1992)
<b>Average No. of Years to Breakthrough:</b>		<b>3 years</b>	<b>18 years</b>

**Собирайте данные, а не изобретайте алгоритмы**

<https://www.kdnuggets.com/2016/05/datasets-over-algorithms.html>

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

## Fashion-MNIST

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



<https://github.com/zalandoresearch/fashion-mnist>

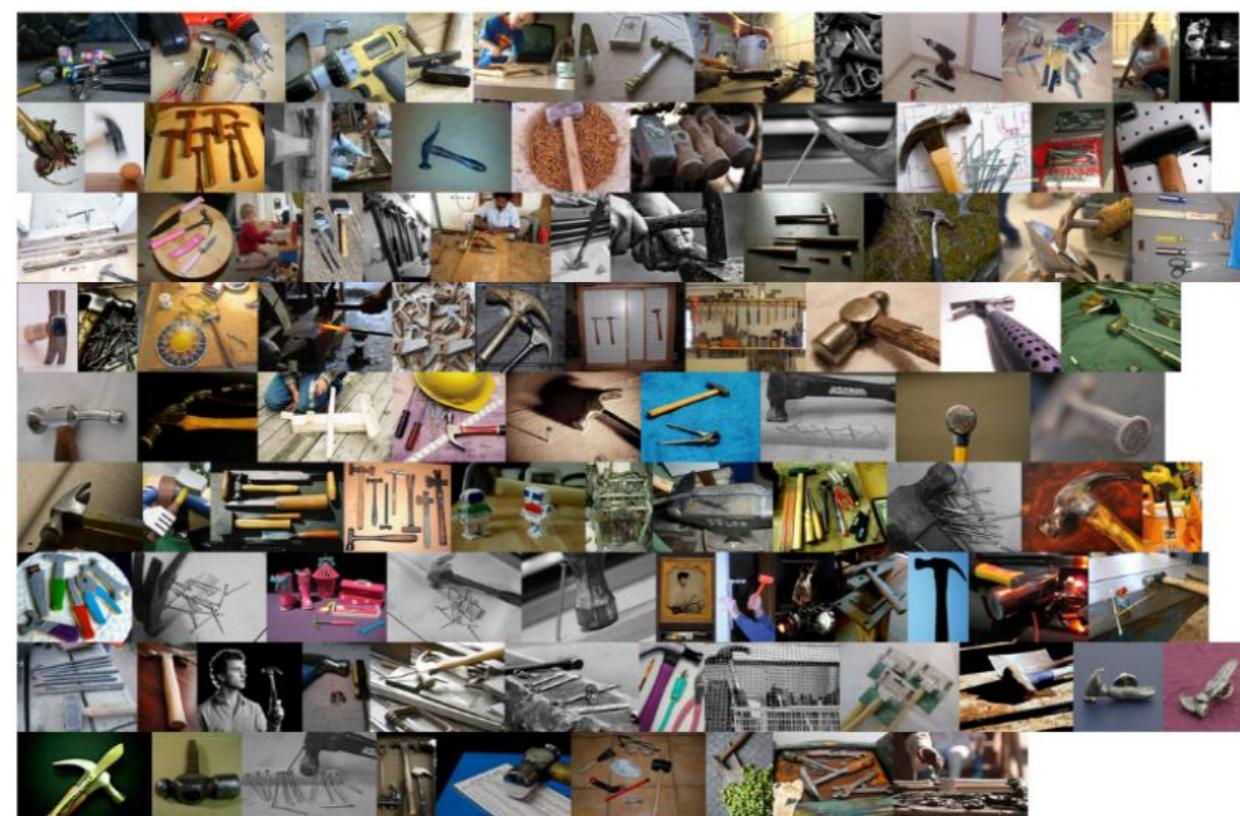
## ImageNet



**1.2 млн. 225x 225-изображений, 1000 классов, метки – MTurk**

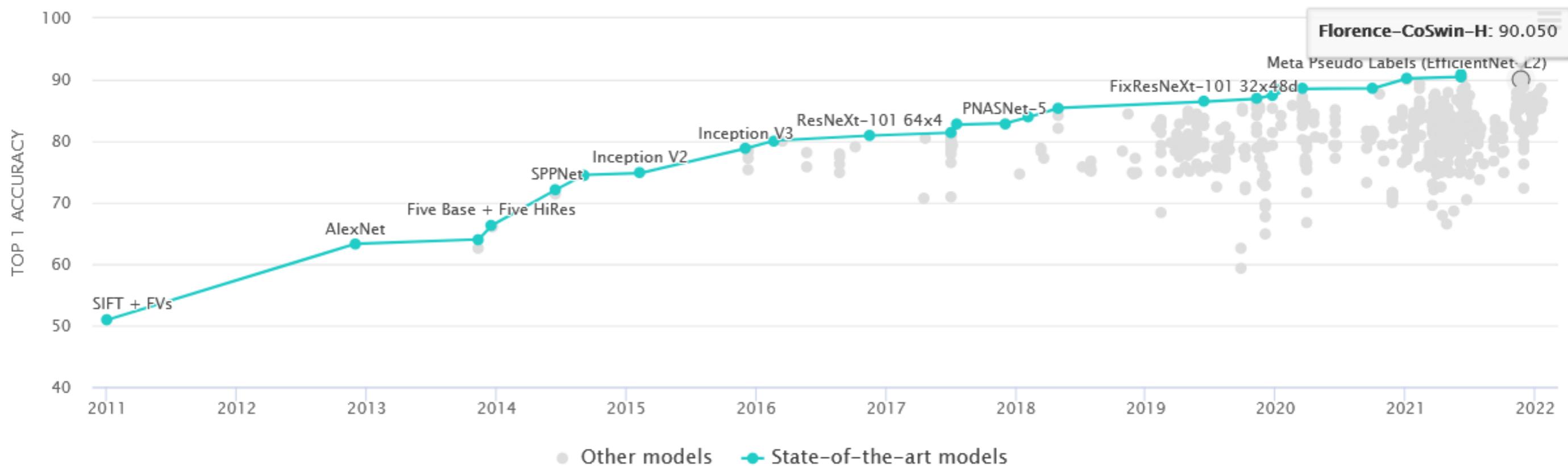
**есть 12 млн. 21000 классов**

### Класс «Молотки»



<http://image-net.org/challenges/LSVRC/2014/browse-synsets>

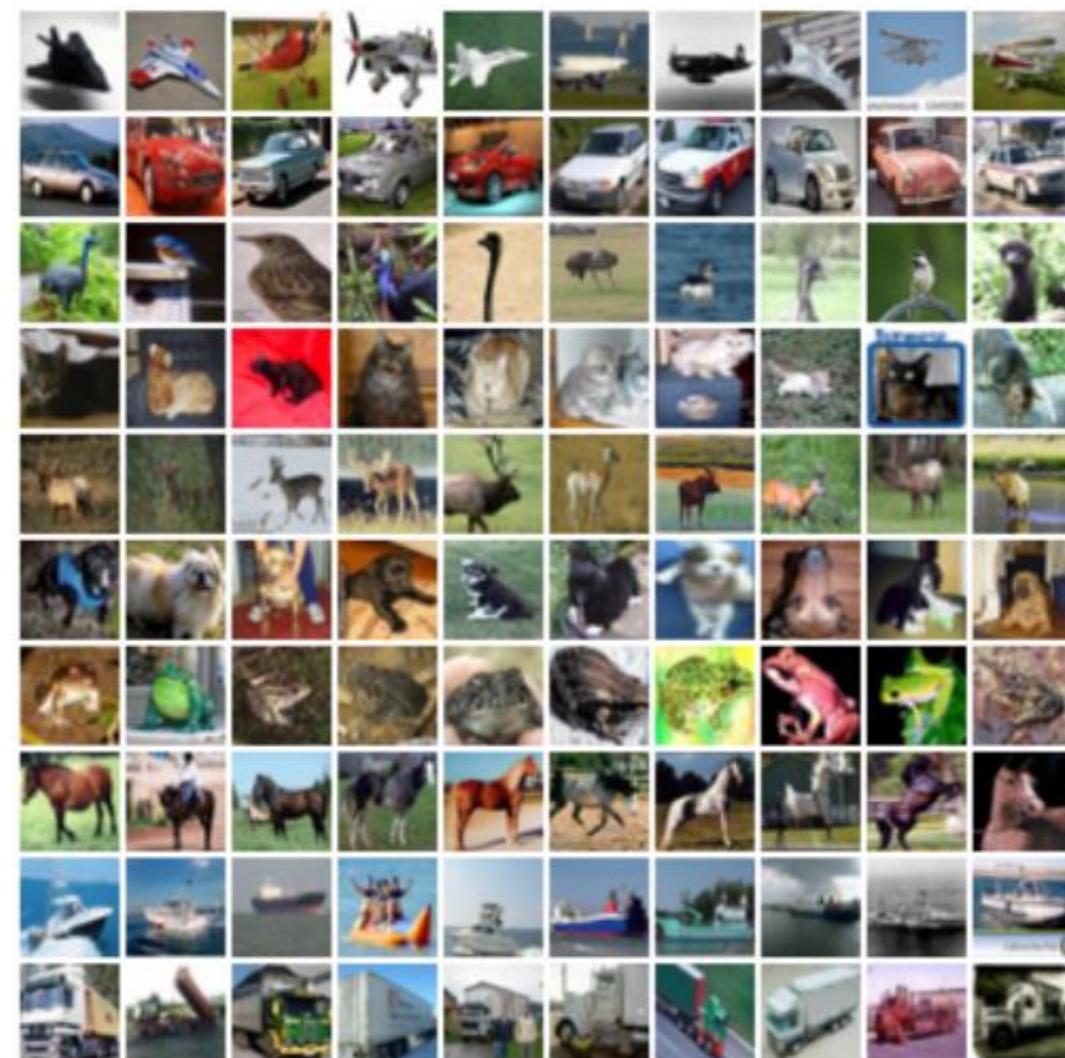
## ImageNet



<https://paperswithcode.com/sota/image-classification-on-imagenet>

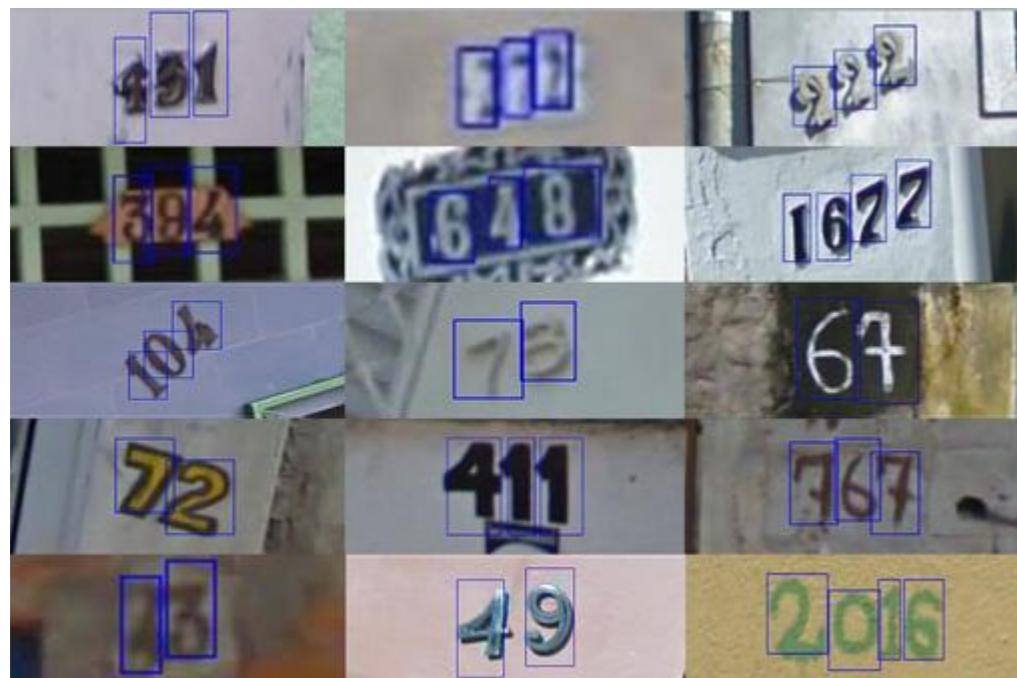
## CIFAR10

**32×32 цветные, 50k – обучение, 10k – тест, 10 классов:**  
**самолет, автомобиль, птица, кошка, олень, собака, лягушка, лошадь, корабль и грузовик**



## The Street View House Numbers (SVHN) Dataset

- 10 classes, 1 for each digit
- 73257 = training, 26032 = testing, and 531131 = additional
  - 2 formats:



Оригинальные картинки с регионами цифр

32×32 – как в MNIST вырезанные регионы

<http://ufldl.stanford.edu/housenumbers/>

<https://paperswithcode.com/sota/image-classification-on-svhn>

## Places

places



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

**10M изображений, 400+ классов**

bedroom



mountain

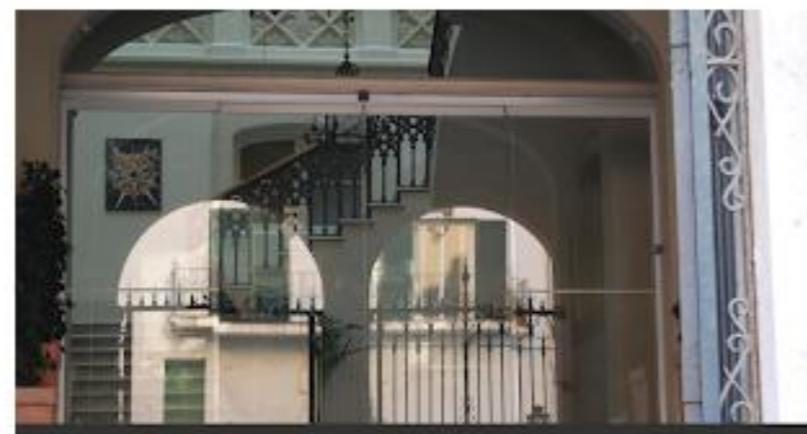


## Open Images Dataset

**object location annotations**

**9М изображений, 6000 классов,**

**16М регионов для 600 классов объектов на 1.9М изображениях**



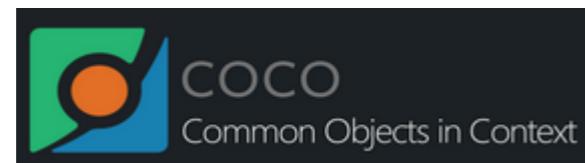
balcony, stairs, facade, iron, door, interior design, gate, architecture, handrail, baluster, window, arch



cutlery, tableware, metal, tool, spoon, fork

<https://github.com/openimages/dataset>

## Microsoft COCO



<http://cocodataset.org/#home>

- **Object segmentation**
- **Recognition in context**
- **Superpixel stuff segmentation**
- **330K images (>200K labeled)**
- **1.5 million object instances**
- **80 object categories**
- **91 stuff categories**
- **5 captions per image**
- **250,000 people with keypoints**



## Изображения и видео – датасеты

Summary of commonly used image and video datasets. Note that image datasets can be used to learn image features, while video datasets can be used to learn both image and video features.

Dataset	Data Type	Size	Synthetic	# classes	Label
ImageNet [13]	Image	1.3 million images	✗	1,000	Object category label
Places [107]	Image	2.5 million images	✗	205	scene categories label
Places365 [108]	Image	10 million images	✗	434	scene categories label
SUNCG [109]	Image	150,000 images	✓	84	depth, volumetric data
MNIST [110]	Image	70,000 images	✗	10	Digit class label
SVHN [111]	Image	600,000 Images	✗	10	Digit class label
CIFAR10 [112]	Image	60,000 Images	✗	10	Object category label
STL-10 [113]	Image	101,300 Images	✗	10	Object category label
PASCAL VOC [96]	Image	2,913 images	✗	20	Category label, bounding box, segmentation mask
YFCC100M [114]	Image/Video	100 million media data	✗	—	Hashtags
SceneNet RGB-D [115]	Video	5 million images	✓	13	Depth, Instance Segmentation, Optical Flow
Moment-in-Time [116]	Video	1 million 3-second videos	✗	339	Video category class
Kinetics [17]	Video	0.5 million 10-second videos	✗	600	Human action class
AudioSet [117]	Video	2 million 10-second videos	✗	632	Audio event class
KITTI [118]	Video	28 videos	✗	—	Data captured by various sensors are available
UCF101 [63]	Video	10,031 videos	✗	101	Human action class
HMDB51 [106]	Video	6,766 videos	✗	51	Human action class

<https://arxiv.org/pdf/1902.06162.pdf>

## Cityscapes Dataset

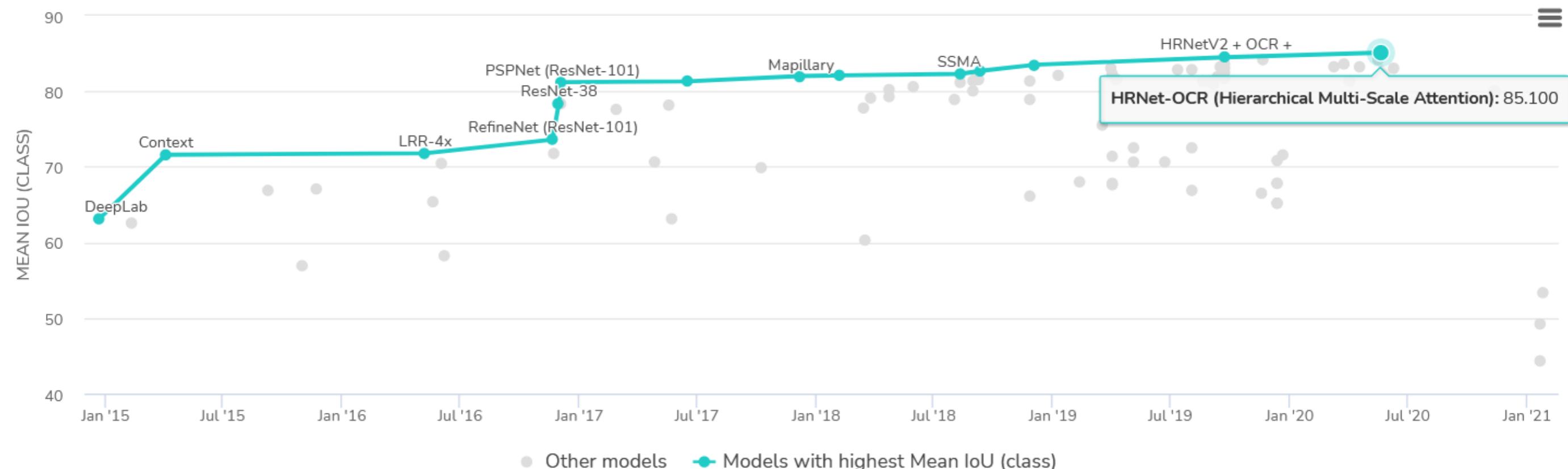
Аннотированные изображения



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

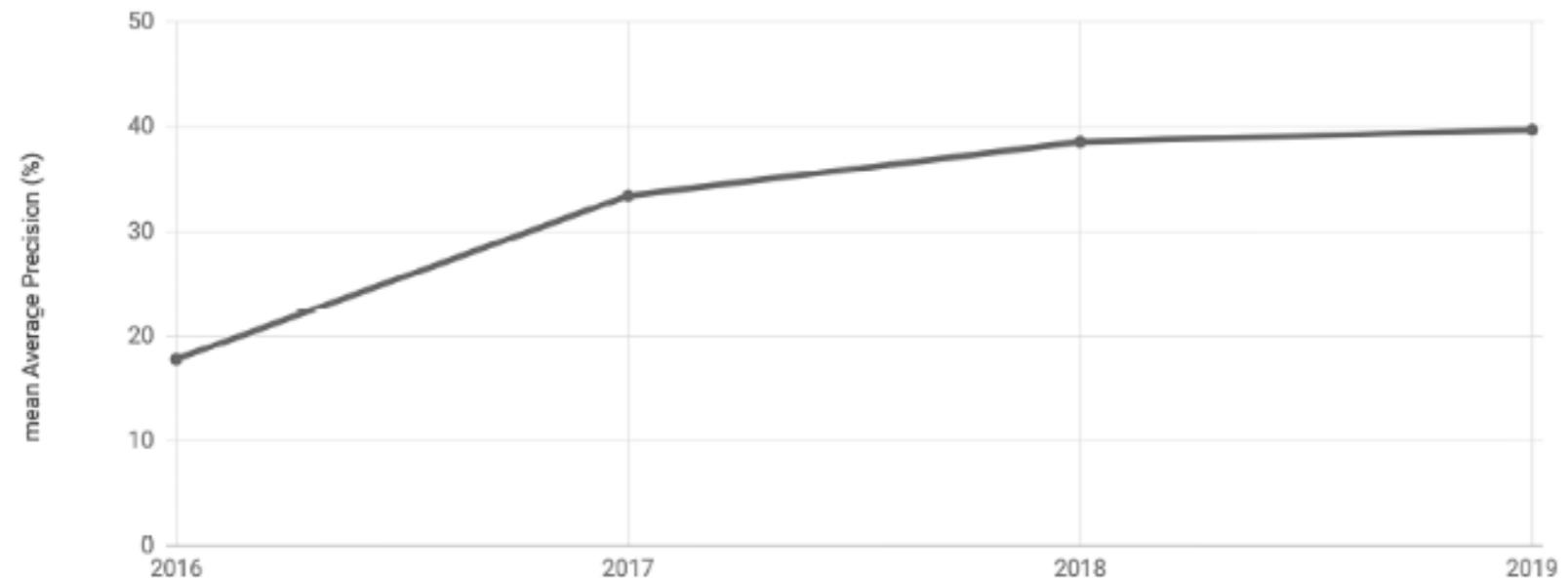
## Cityscapes Dataset

### Семантическая сегментация



<https://paperswithcode.com/sota/semantic-segmentation-on-cityscapes>

## ActivityNet



### A Large-Scale Video Benchmark for Human Activity Understanding

<http://activity-net.org>

**200 activity classes**

- **10,024 training videos (15,410 instances)**
- **4,926 validation videos (7,654 instances)**
- **5,044 testing videos (labels withheld)**

# GLUE (General Language Understanding Evaluation) / superGLUE

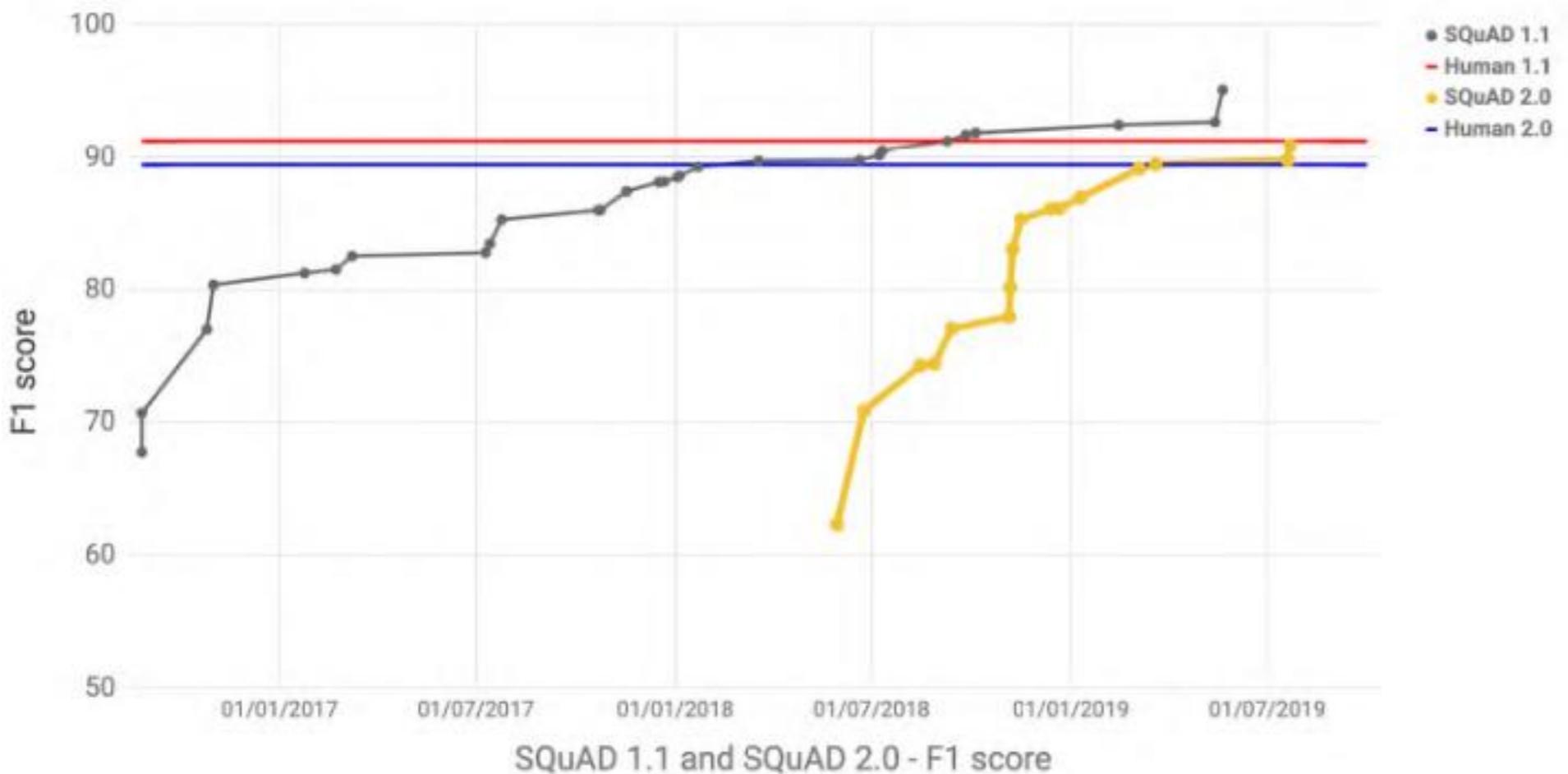
<https://gluebenchmark.com/leaderboard>

Rank	Name	Model	URL	Score	CoLA	SST-2	MRPC	STS-B	QQP	MNLI-m	MNLI-mm	QNLI	RTE	WNLI	AX	
1	DeBERTa Team - Microsoft	DeBERTa / TuringNLVRv4		90.8	71.5	97.5	94.0/92.0	92.9/92.6	76.2/90.8	91.9	91.6	99.2	93.2	94.5	53.2	
2	HFL iFLYTEK	MacALBERT + DKM		90.7	74.8	97.0	94.5/92.6	92.8/92.6	74.7/90.6	91.3	91.1	97.8	92.0	94.5	52.6	
+	3	Alibaba DAMO NLP	StructBERT + TAPT		90.6	75.3	97.3	93.9/91.9	93.2/92.7	74.8/91.0	90.9	90.7	97.4	91.2	94.5	49.1
+	4	PING-AN Omni-Sinitic	ALBERT + DAAF + NAS		90.6	73.5	97.2	94.0/92.0	93.0/92.4	76.1/91.0	91.6	91.3	97.5	97.5	4.5	51.2
5	ERNIE Team - Baidu	ERNIE		90.4	74.4	97.5	93.5/91.4	93.0/92.6	75.2/90.9	91.4	91.0	96.6	90.9	94.5	51.7	
15	GLUE Human Baselines	GLUE Human Baselines		87.1	66.4	97.8	86.3/80.8	92.7/92.6	59.5/80.4	92.0	92.8	91.2	93.6	95.9	-	

<https://super.gluebenchmark.com/leaderboard>

Rank	Name	Model	URL	Score	BoolQ	CB	COPA	MultiRC	ReCoRD	RTE	WiC	WSC	AX-g	AX-b	
+	1	DeBERTa Team - Microsoft	DeBERTa / TuringNLVRv4		90.3	90.4	95.7/97.6	98.4	88.2/63.7	94.5/94.1	93.2	77.5	95.9	93.3/93.8	66.7
+	2	Zirui Wang	T5 + Meena, Single Model (Meena Team - Google Brain)		90.2	91.3	95.8/97.6	97.4	88.3/63.0	94.2/93.5	92.7	77.9	95.9	88.8/89.9	66.5
+	3	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	99.3/99.7	76.6
+	4	T5 Team - Google	T5		89.3	91.2	93.9/96.8	94.8	88.1/63.3	94.1/93.4	92.5	76.9	93.8	92.7/91.9	65.6
+	5	Huawei Noah's Ark Lab	NEZHA-Plus		86.7	87.8	94.4/96.0	93.6	84.6/55.1	90.1/89.6	89.1	74.6	93.2	87.1/74.4	58.0

## Stanford Question Answering Dataset (SQuAD)



In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under **gravity**. The main forms of precipitation include drizzle, rain, sleet, snow, **graupel** and hail... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals **within a cloud**. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?  
**gravity**

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?  
**graupel**

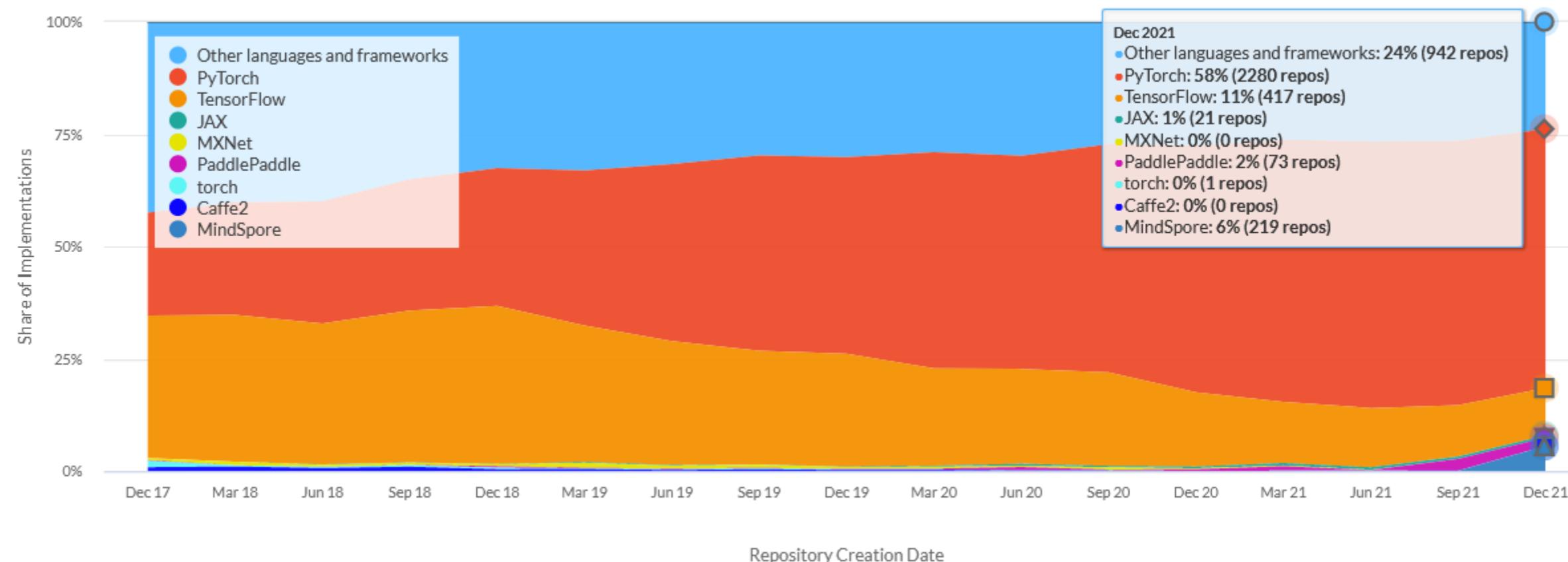
Where do water droplets collide with ice crystals to form precipitation?  
**within a cloud**

**Figure 1:** Question-answer pairs for a sample passage in the SQuAD dataset. Each of the answers is a segment of text from the passage.

**вопросы по статье вики, ответ – фрагмент**

<https://arxiv.org/abs/1606.05250>

## Фреймворки



<https://paperswithcode.com/trends>

## Документация

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

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

## Пример: 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

## Книги

**лучшая книга: Ian Goodfellow, Yoshua Bengio, Aaron Courville «Deep Learning»**

<http://www.deeplearningbook.org/>

**Хорошая обзорная, но примеры на Keras**

**Mohamed Elgendi «Deep Learning for Vision Systems»**

## Датасеты

<https://www.datasetlist.com/>

**Текущие рекорды (но уже не обновляется)**

<https://paperswithcode.com>

<https://www.stateofheart.ai>

<https://sotabench.com>

<https://www.eff.org/ai/metrics>

## Пограться с NN

<https://vk.com/@etorabotaet-21-sait-gde-mozhno-protestirovat-rabotu-neirosetei>

## Другие курсы

**Классика DL CS231n: «Convolutional Neural Networks for Visual Recognition from Stanford»**

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

**Антон Осокин Материалы курса «Глубинное обучение»// ФКН ВШЭ, 2018 (есть 2020 год)**

[https://github.com/aosokin/DL\\_CSHSE\\_spring2018](https://github.com/aosokin/DL_CSHSE_spring2018)

**Хороший обзорный курс CS 598 LAZ: Cutting-Edge Trends in Deep Learning and Recognition**

<http://slazebni.cs.illinois.edu/spring17/>

**Отличный курс CS294-158-SP20 «Deep Unsupervised Learning»**

<https://sites.google.com/view/berkeley-cs294-158-sp20/home>

**Для начинающих по DL: «Методы обработки и анализа больших данных» (Роман Соловьёв)**

<https://youtube.com/playlist?list=PLMyPRULSXkaWaJrQOWCBw0nVVYdPsPPj>

**Хорошая практика «Deep Learning»**

<https://atcold.github.io/pytorch-Deep-Learning/>