курс «Глубокое обучение»

Введение

Александр Дьяконов

(ВМК МГУ имени М.В. Ломоносова)

10 февраля 2020 года

Глубокое обучение

мини-курс (?) для студентов 3 курса ММП ВМК МГУ – читается с весны 2019 года

лекции со слайдами
семинары ~ практикум

понедельник, 12:50 (?), 582
Отчётность: доклады + отчёты

Лектор: Дьяконов Александр Геннадьевич
Страница курса: https://github.com/Dyakonov/DL/

телеграм-чат (не пишите в личку)
https://classroom.google.com

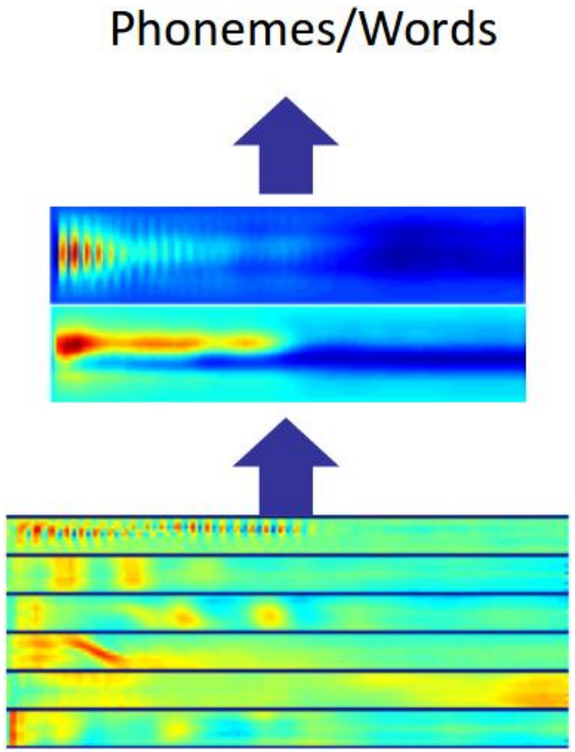
Другие курсы

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

DL: первые успехи

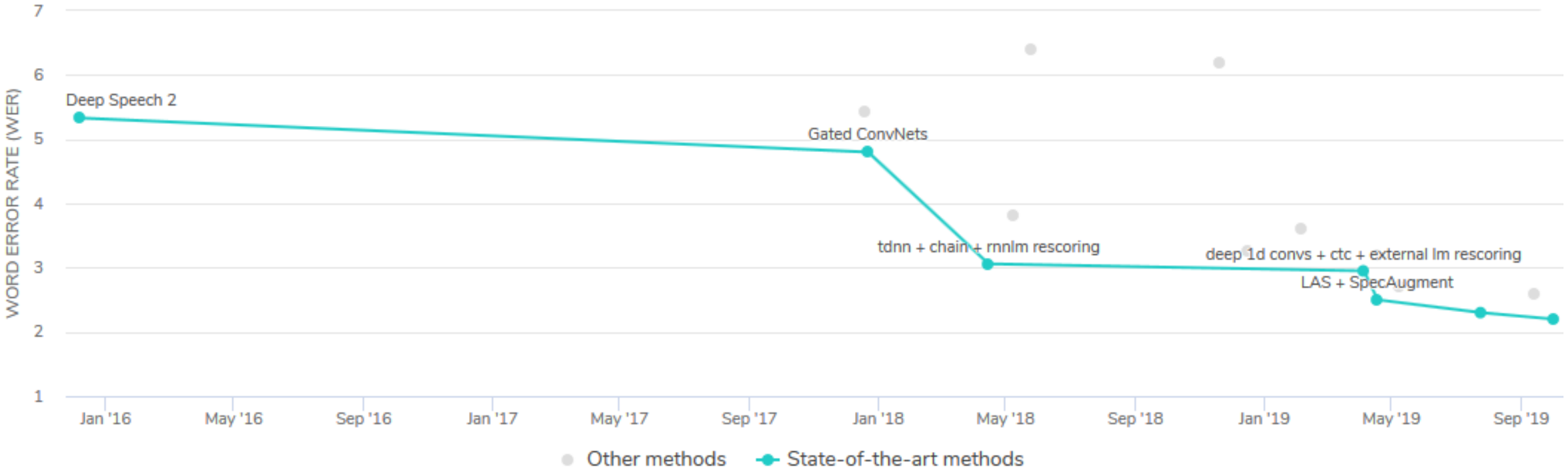
Первый прорыв – речь

Context-Dependent Pre-trained Deep Neural Networks for Large Vocabulary Speech Recognition. Dahl et al. (2010)



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

Распознавание речи



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

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

DL: первые успехи

Results IJCNN2011 COMPETITION

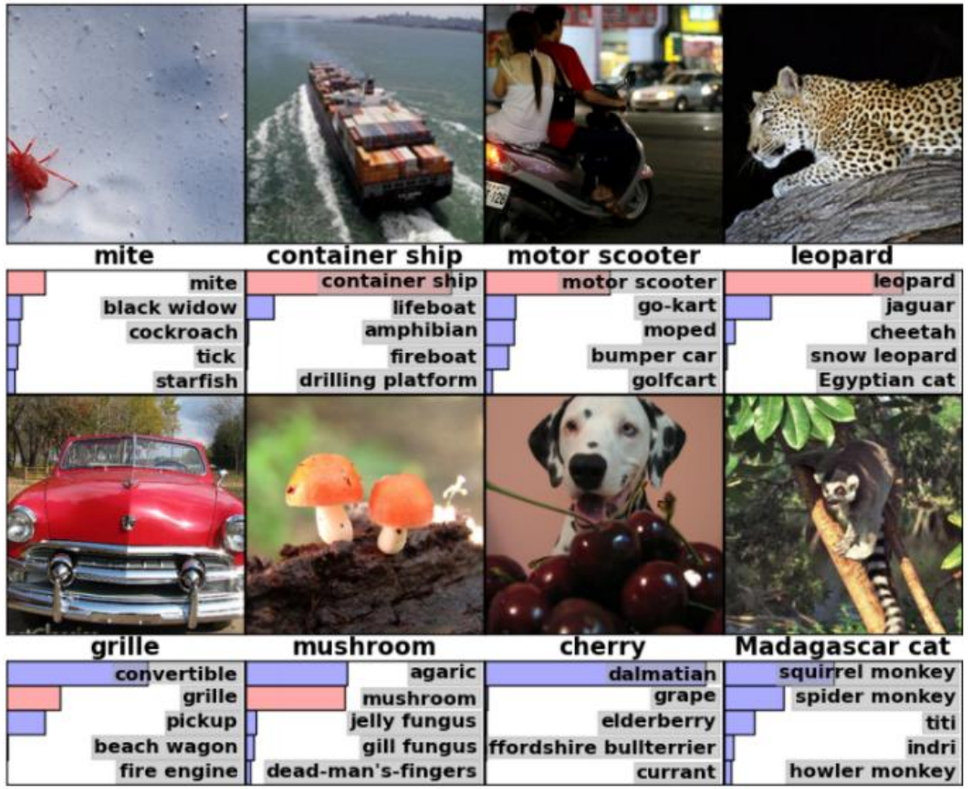
TEAM	METHOD	TOTAL	SUBSET ALL SIGNS
[3] IDSIA	Committee of CNNs	99.46%	99.46%
[1] INI-RTCV	Human Performance	98.84%	98.84%
[4] sermanet	Multi-Scale CNNs	98.31%	98.31%
[2] CAOR	Random Forests	96.14%	96.14%
[6] INI-RTCV	LDA on HOG 2	95.68%	95.68%
[5] INI-RTCV	LDA on HOG 1	93.18%	93.18%
[7] INI-RTCV	LDA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

2011 год

На соревновании по распознаванию дорожных знаков CNN превзошёл точность человека!

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

Классификация изображений



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

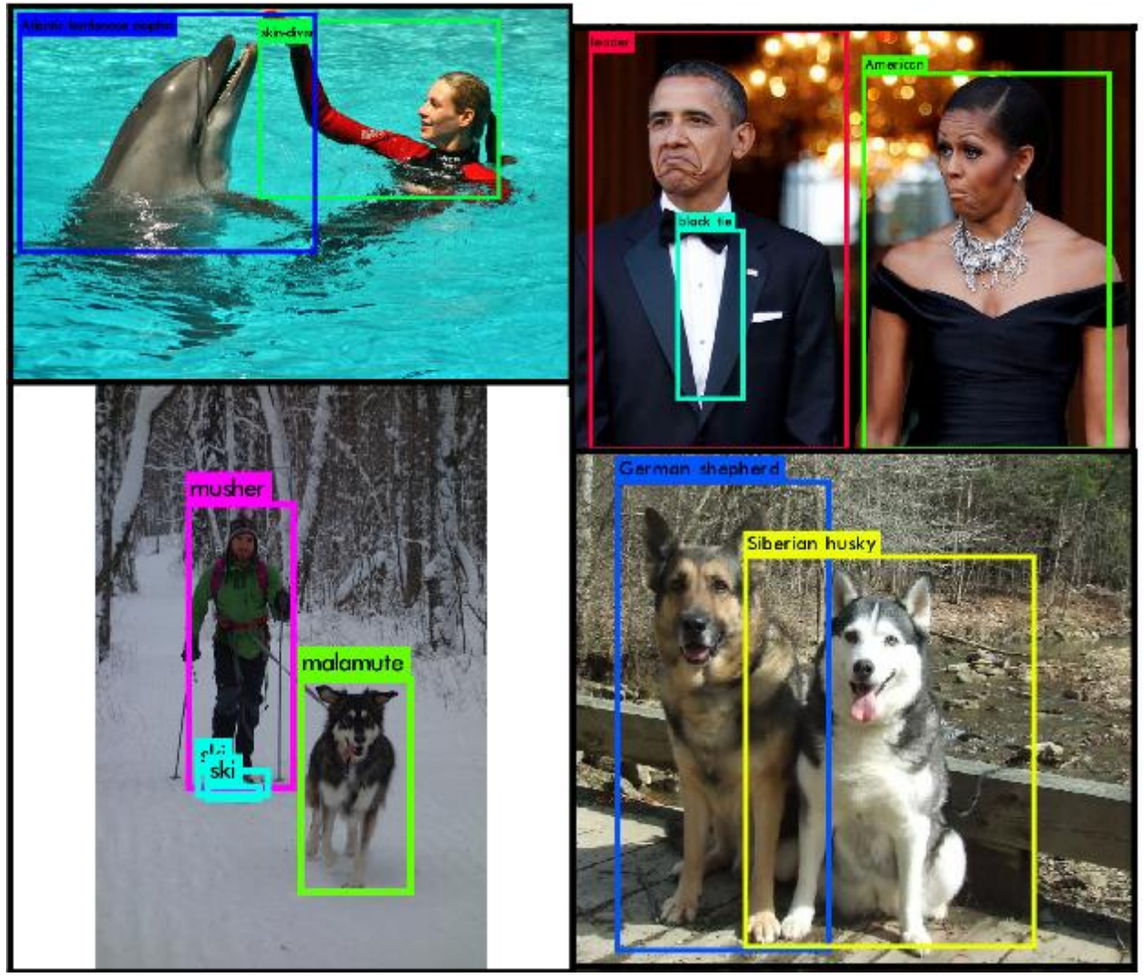
Компьютерное зрение и обработка изображений (Computer Vision / Image Analysis)

Пример Классификации: QuickDraw

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

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)

Распознавание сцен (Scene recognition)

Bedroom

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

Удаление фона (Background removal)

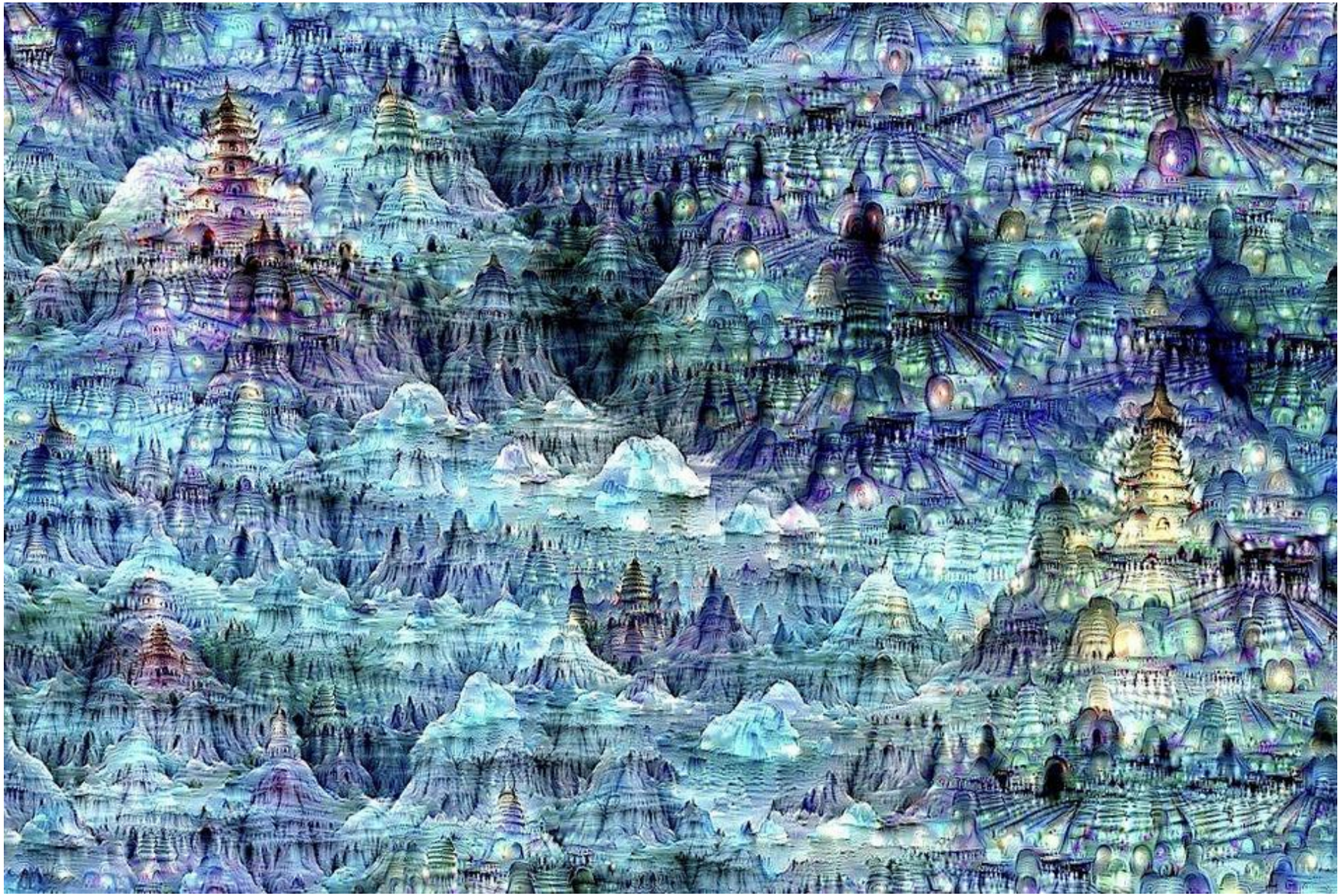
https://towardsdatascience.com/background-removal-with-deep-learning-c4f2104b3157
https://github.com/YunanWu2168/Background-removal-using-deep-learning

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

Определение позы (Human pose estimation)

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

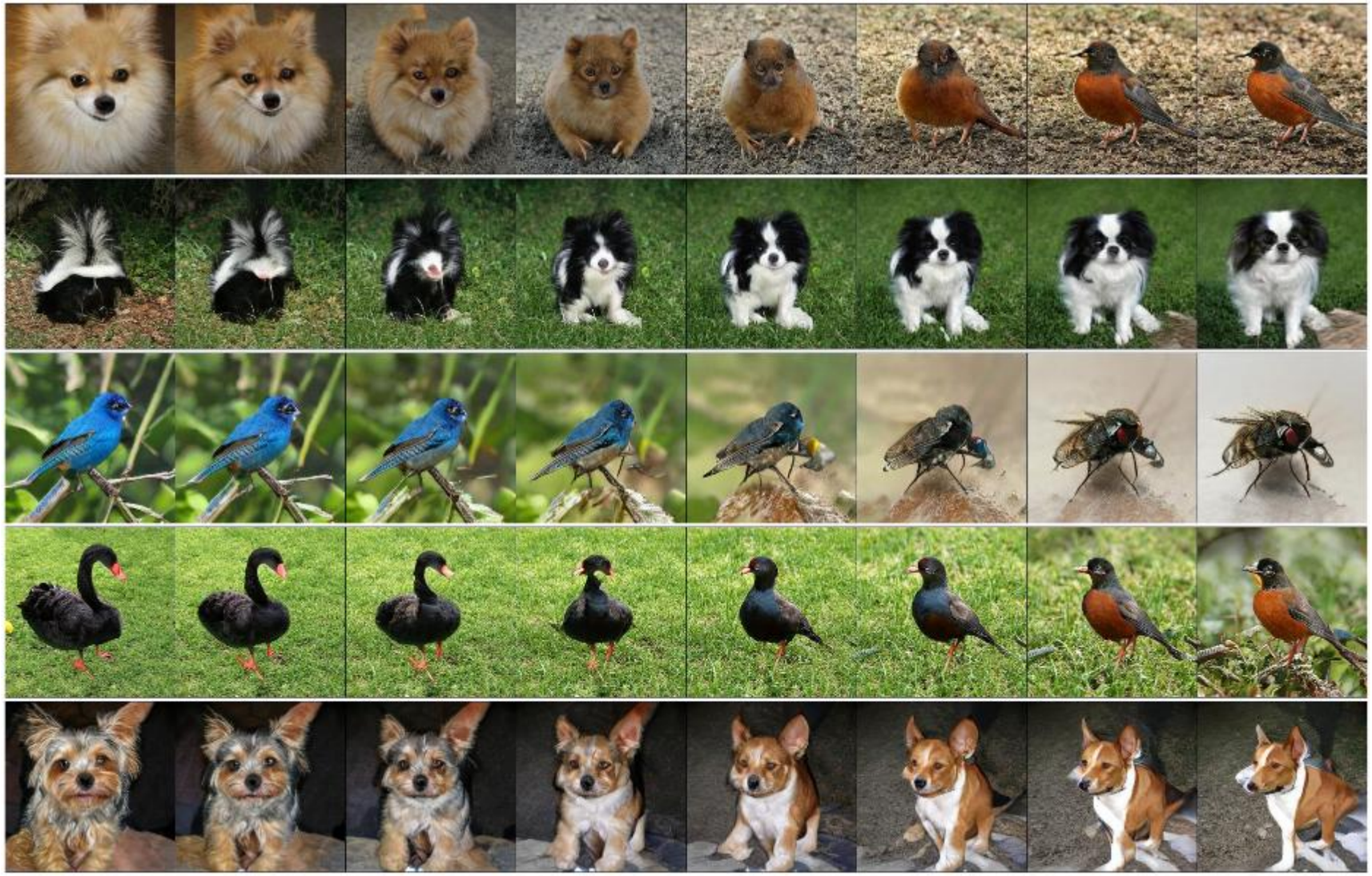
Обработка изображений: генерация изображений



Обработка изображений: генерация изображений

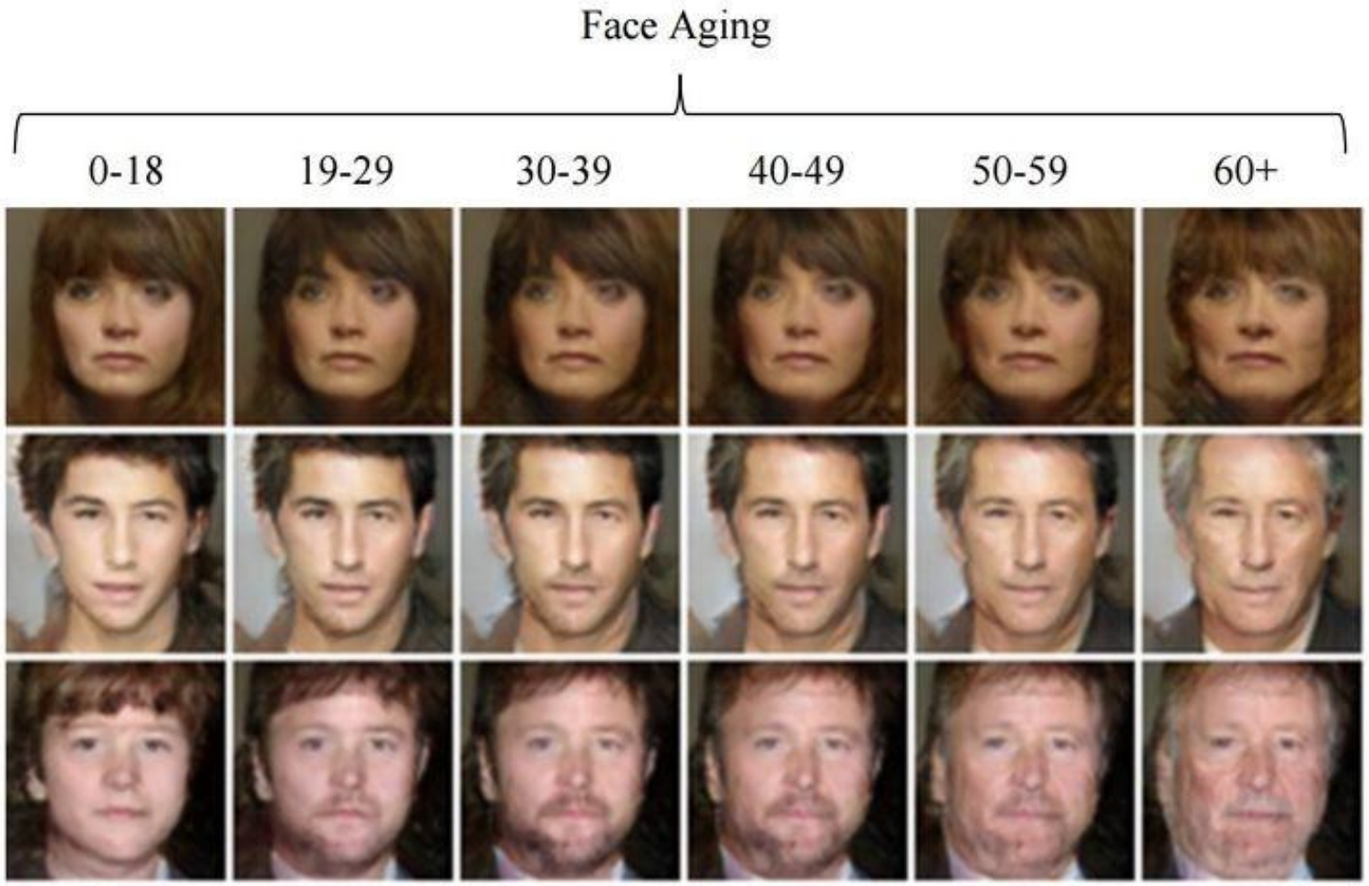
https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf
http://research.nvidia.com/publication/2017-10_Progressive-Growing-of

BigGAN: Генерация изображений / интерполяция



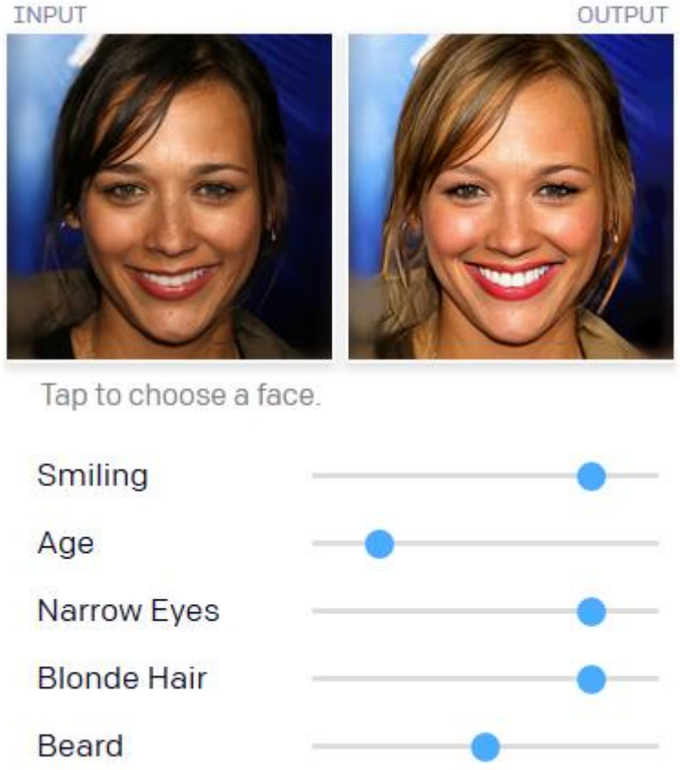
https://arxiv.org/pdf/1809.11096.pdf

Обработка изображений: изменение изображений



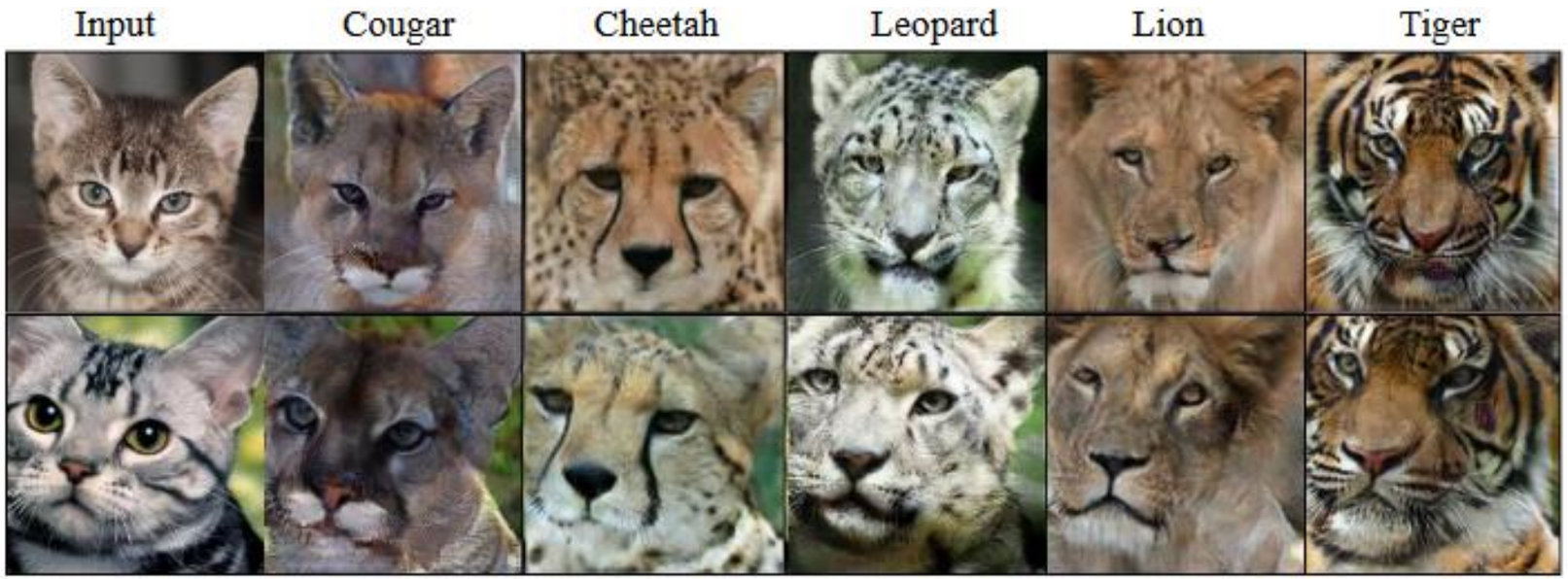
https://arxiv.org/pdf/1702.01983.pdf

Обработка изображений: изменение изображений



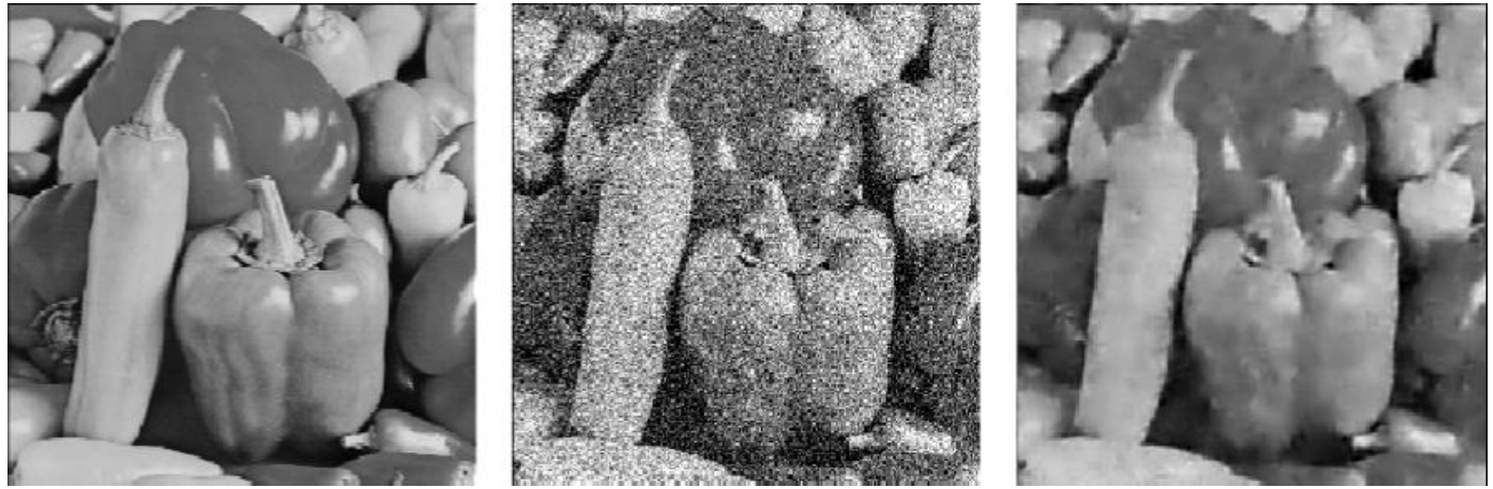
https://openai.com/blog/glow/

Обработка изображений: трансляция изображений (Image-to-Image Translation)

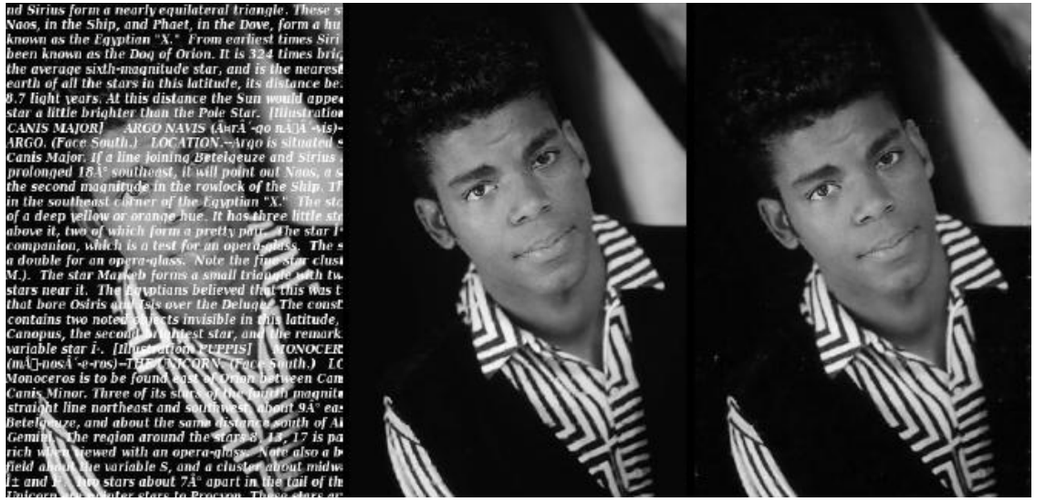


https://arxiv.org/pdf/1703.00848.pdf

Обработка изображений: улучшение изображений: Denoising / Restoring



Inpainting



Обработка изображений: дорисовка изображений (Image completion)

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

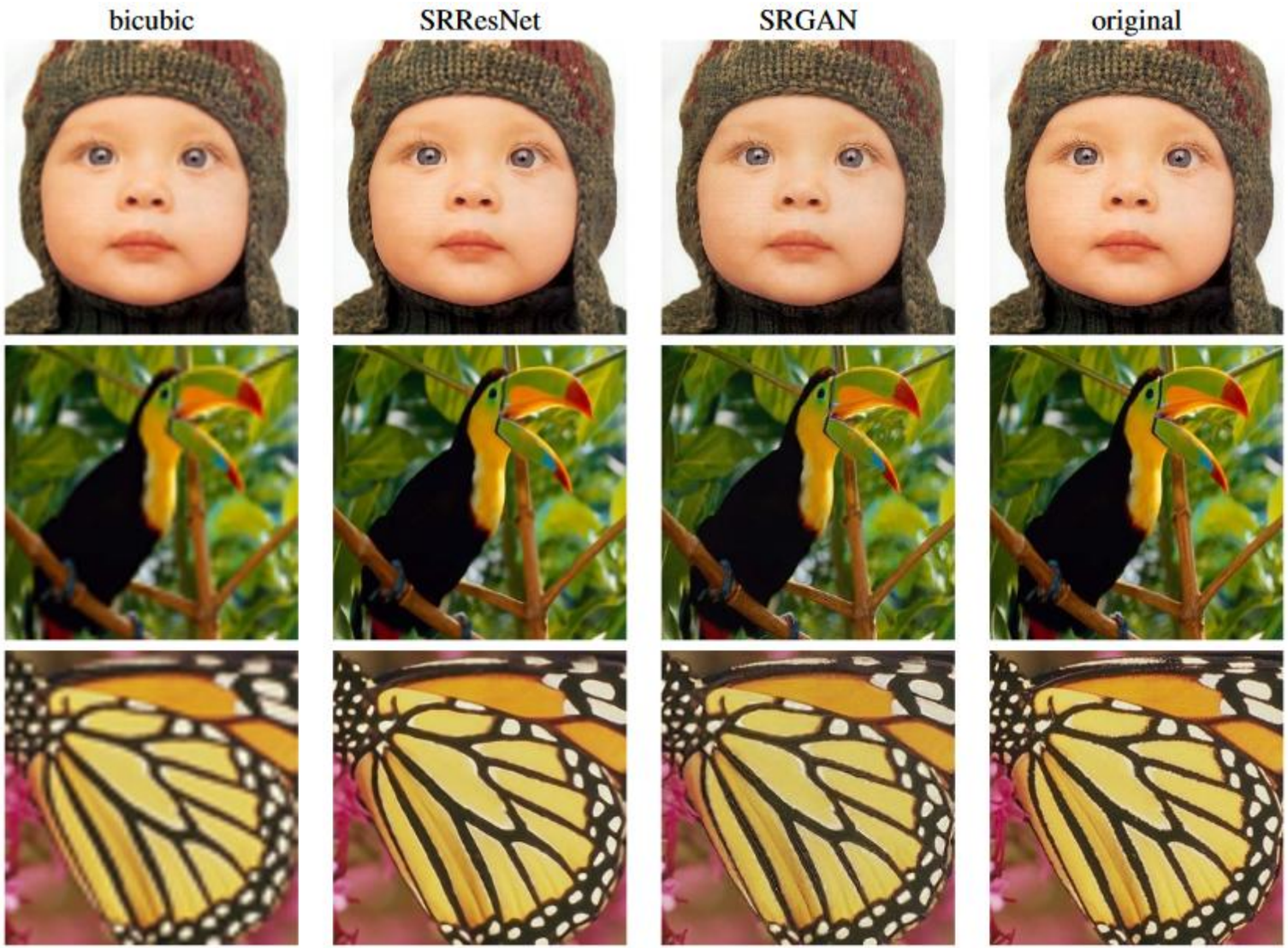
Обработка изображений: стилизация изображений (перенос стиля)

https://deepart.io/

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

(a) Input image (b) Reference style image (e) Our result

Обработка изображений: супер-разрешение (Super-Resolution)



https://arxiv.org/abs/1609.04802

Обработка изображений: улучшение фотографий

в обучении не пары (плохое хорошее),
а примеры хороших и что надо улучшить

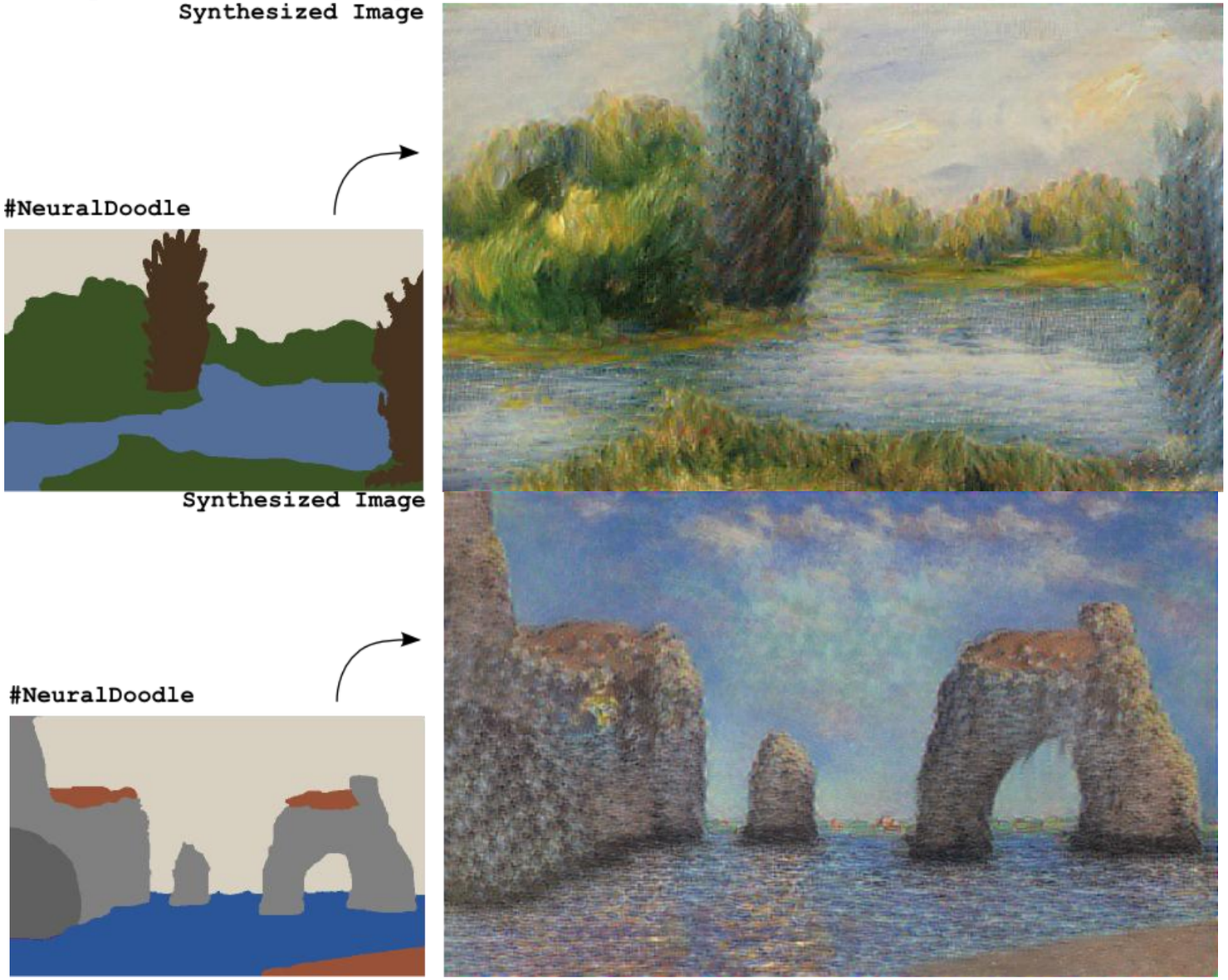
http://www.vision.ee.ethz.ch/~ihnatova/wespe.html

Обработка изображений: раскрашивание изображений

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

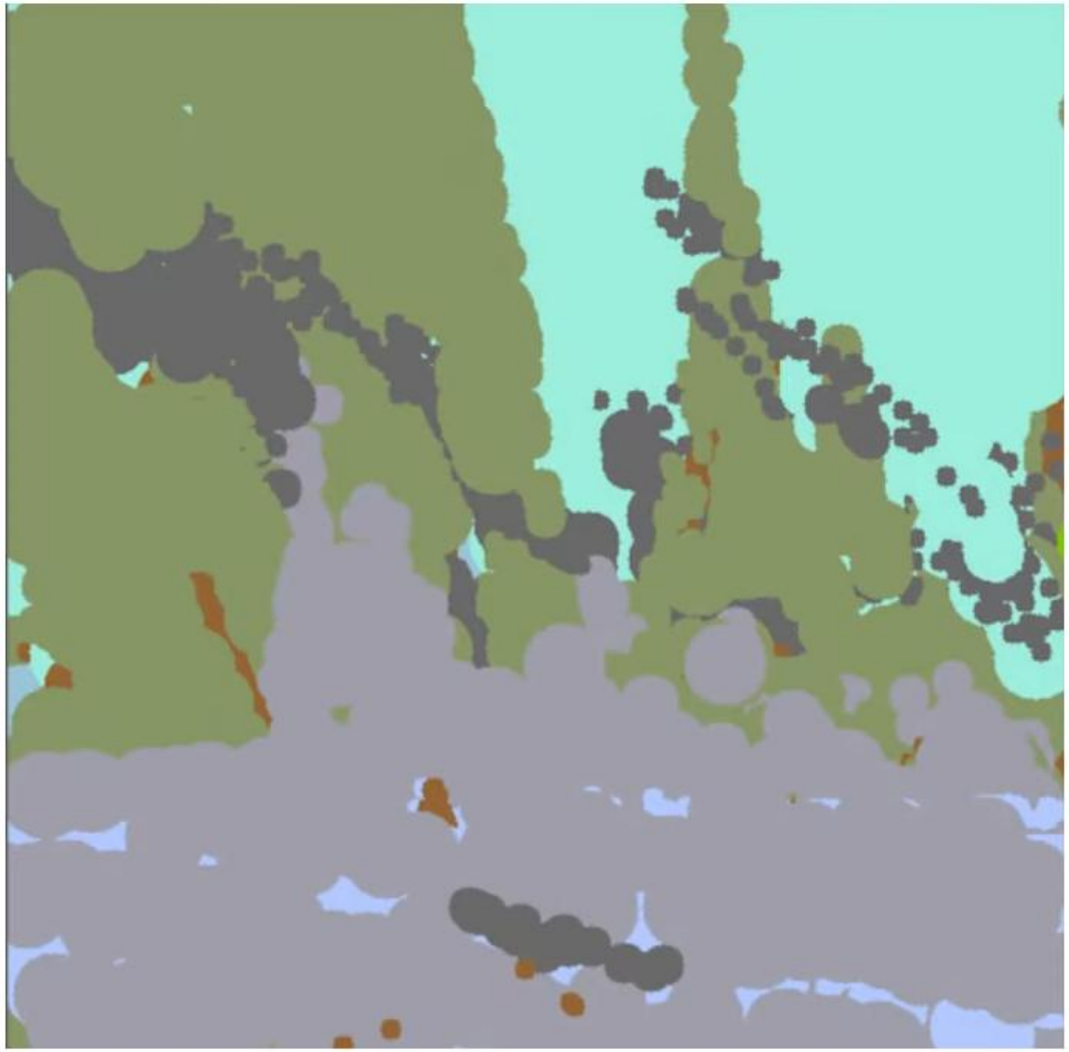
в 20% случаях раскрашенное изображение путали с оригиналом…

Обработка изображений: генерация изображений



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

Обработка изображений: генерация изображений

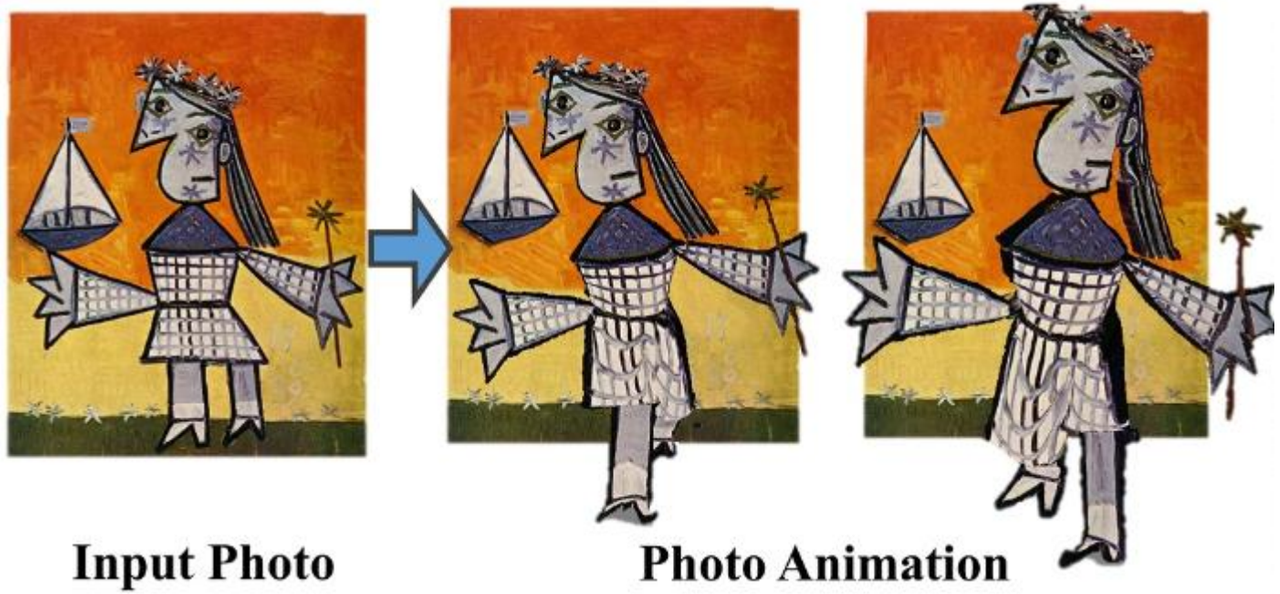


NVIDIA GauGAN Segmentation Input

NVIDIA GauGAN Output

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

Обработка изображений: image → video



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

Задачи для дизайна / креатива: генерация объектов / представление объектов

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

Понимание изображений: аннотирование изображений

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

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

Задачи для сервисов: определение породы собаки по фотографии

Shiba Inu

Lively, independent-minded, aloof toward strangers

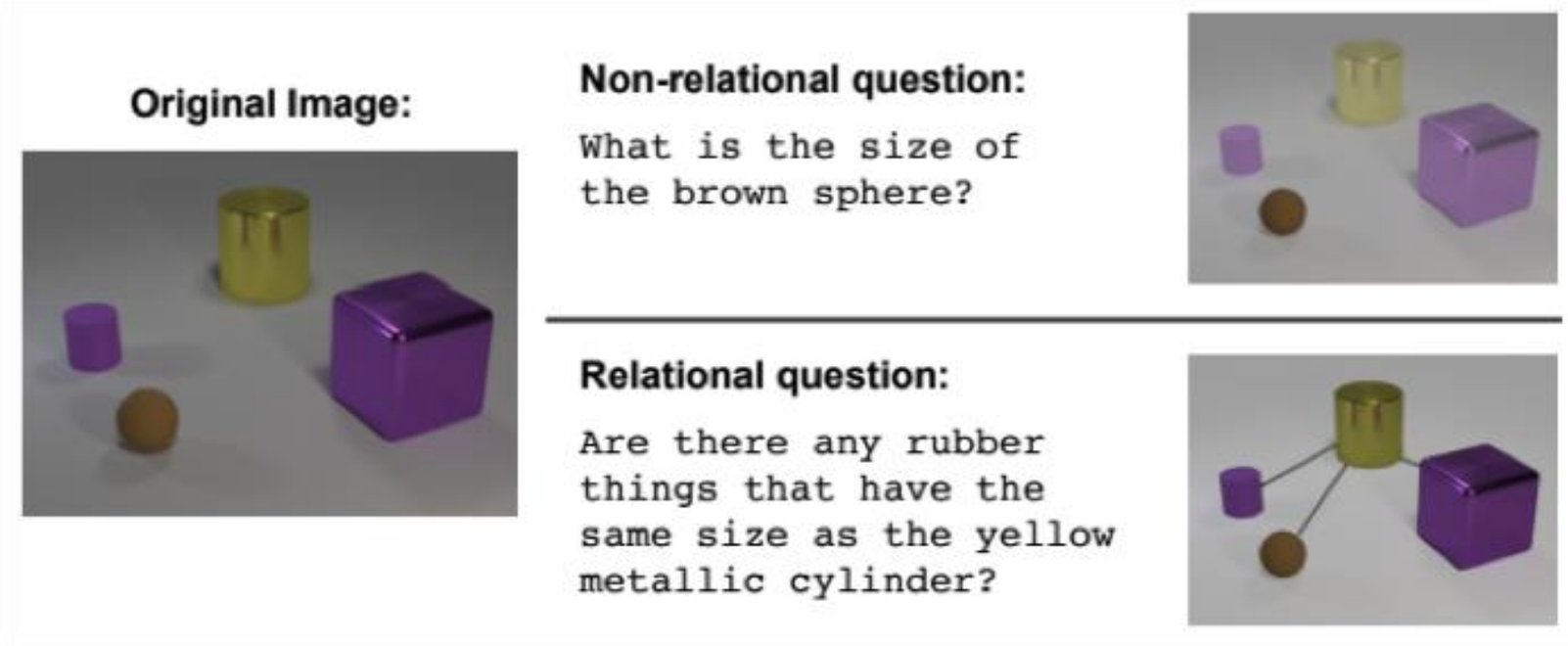
German Shepherd Dog

Must be kept busy, intense, highly intelligent

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

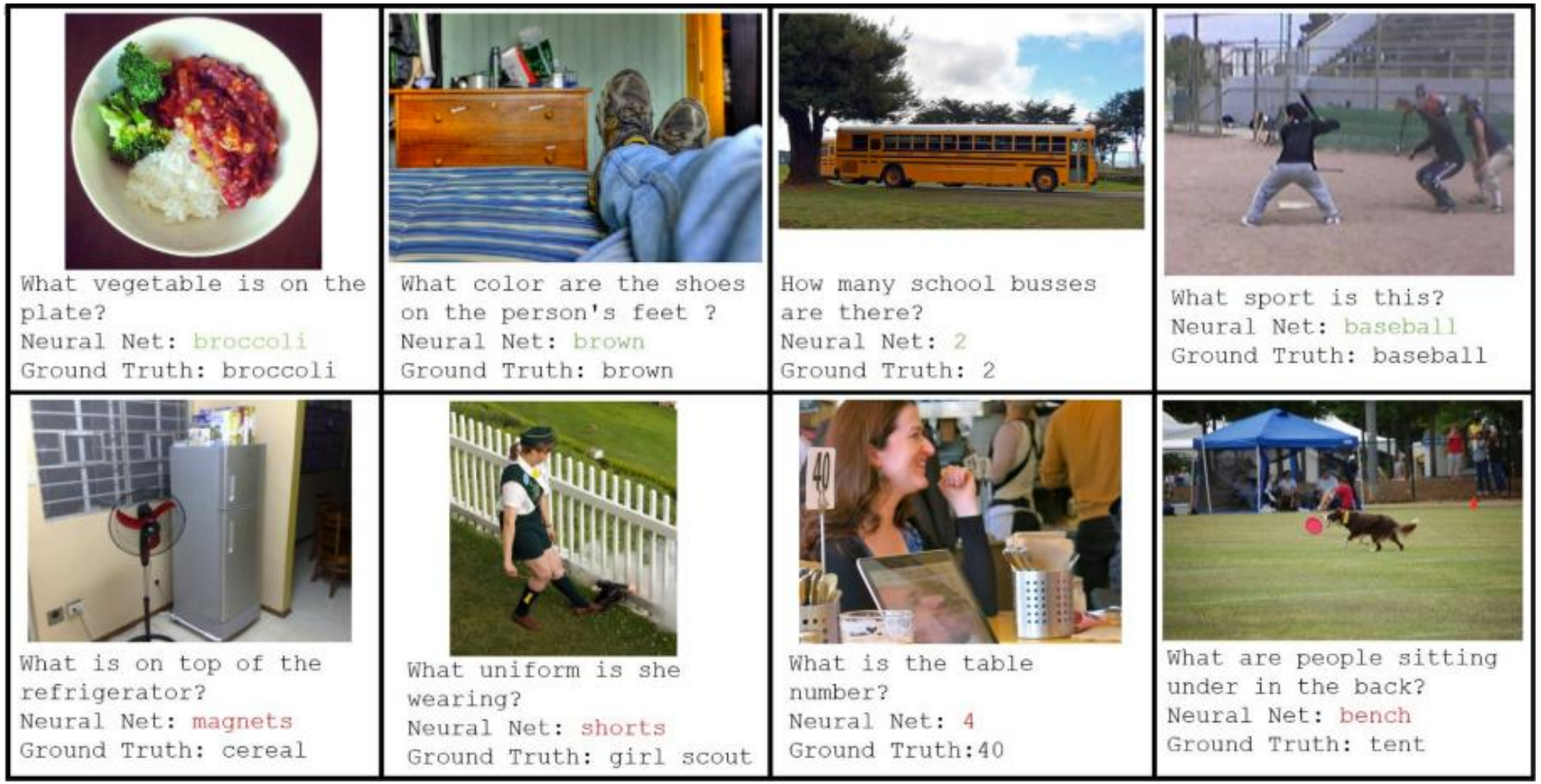
Аналогично: определение марки авто по фото

Понимание изображений: Visual reasoning



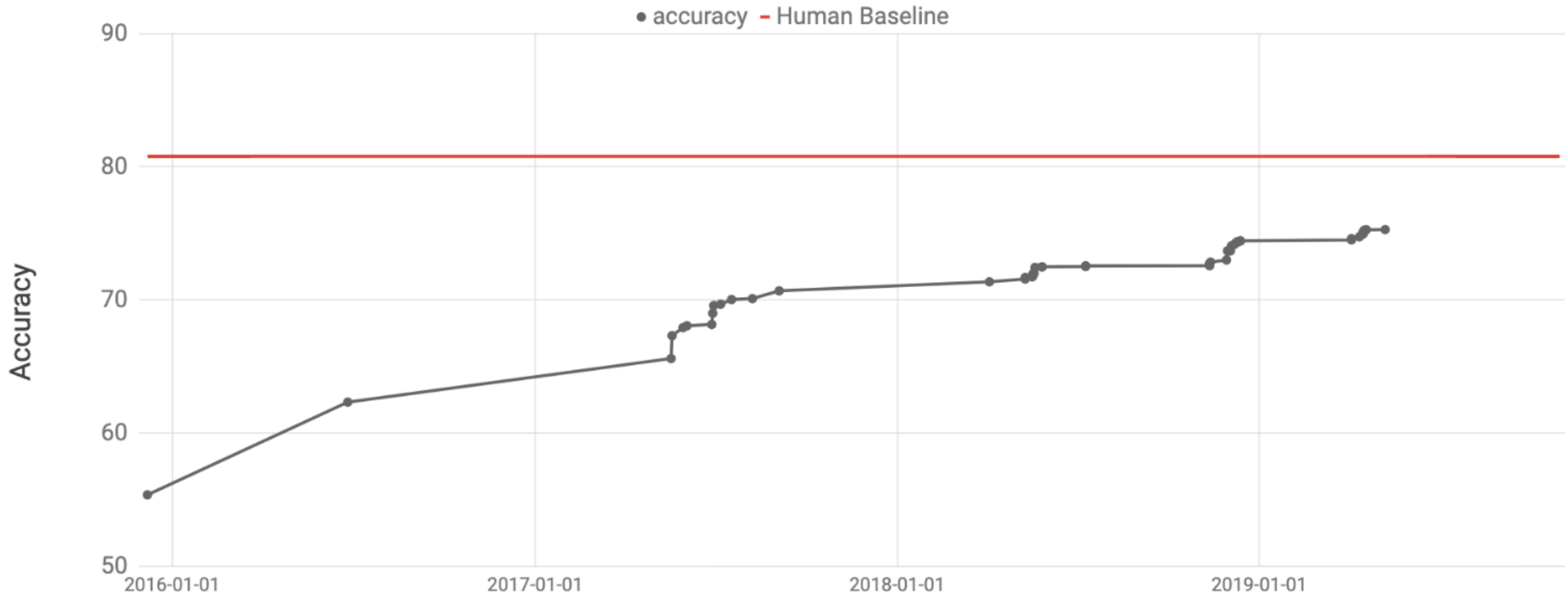
https://arxiv.org/pdf/1706.01427.pdf

Понимание изображений: Visual QA Challenge



Aishwarya Agrawal и др.
«VQA: Visual Question Answering»
https://arxiv.org/abs/1505.00468

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: 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: 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 **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 **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.

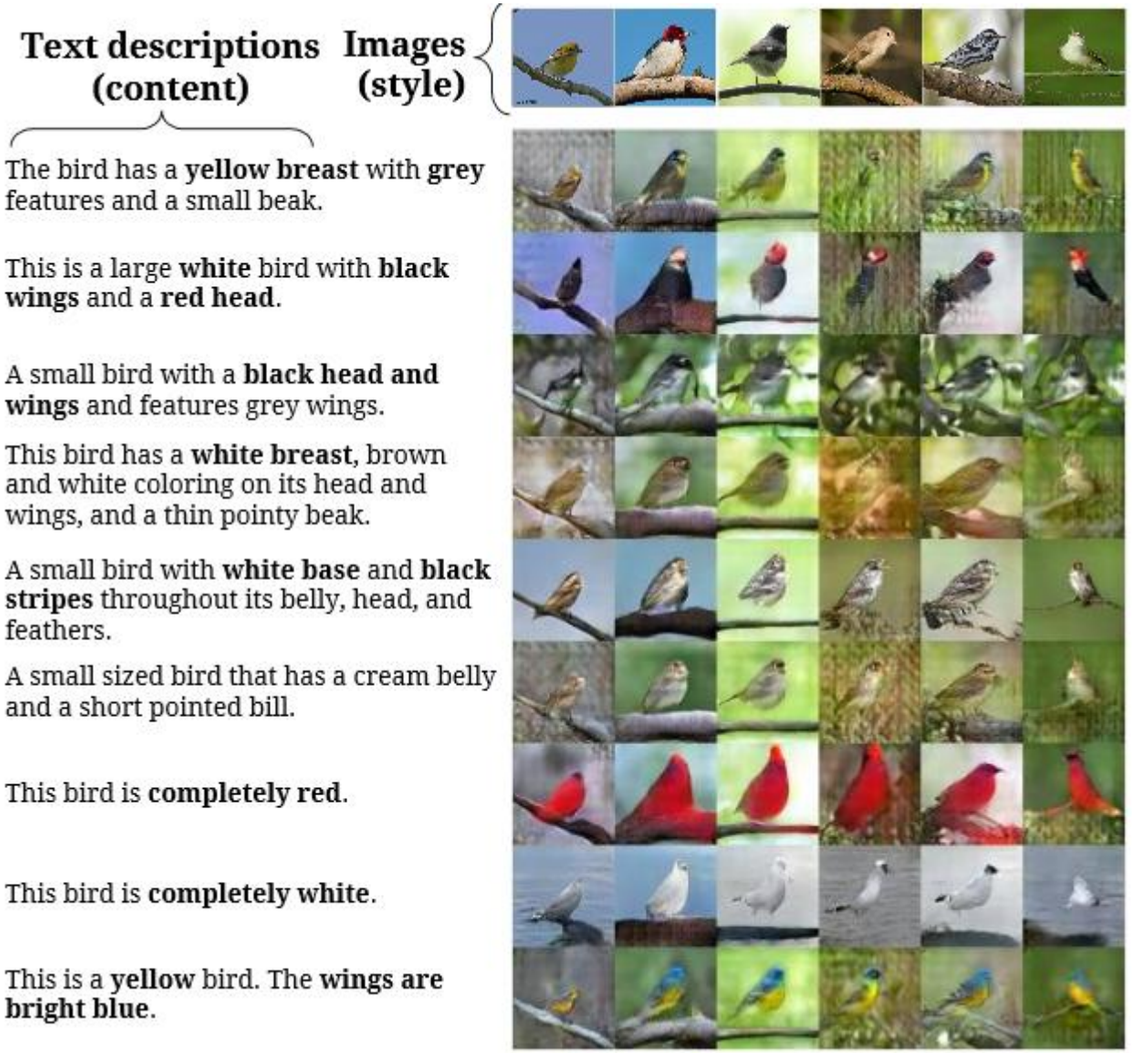
*This is a **Geococcyx** because...*

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.

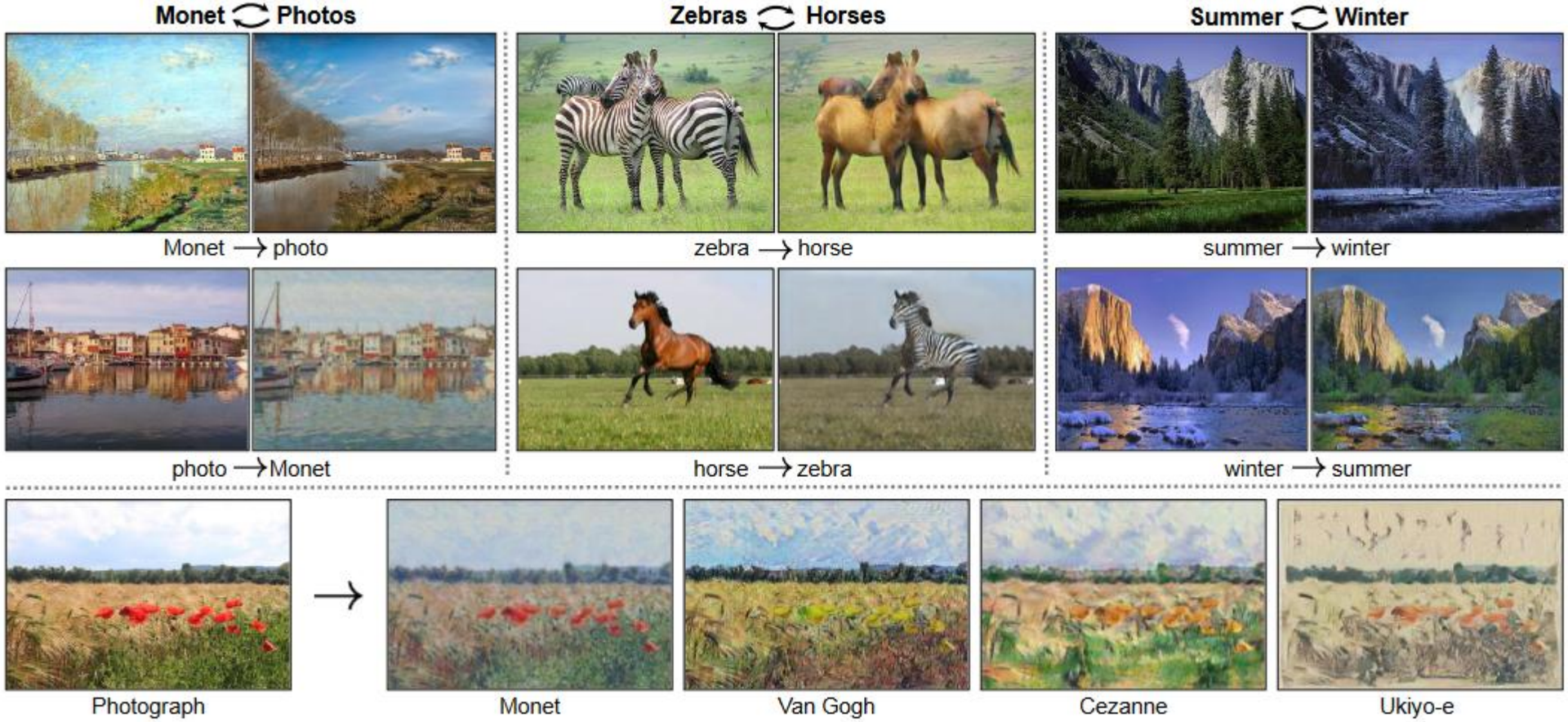
https://arxiv.org/pdf/1603.08507.pdf

Текст в изображение (text2image)



S.Reed «Generative Adversarial Text to Image Synthesis», 2016 https://arxiv.org/pdf/1605.05396.pdf

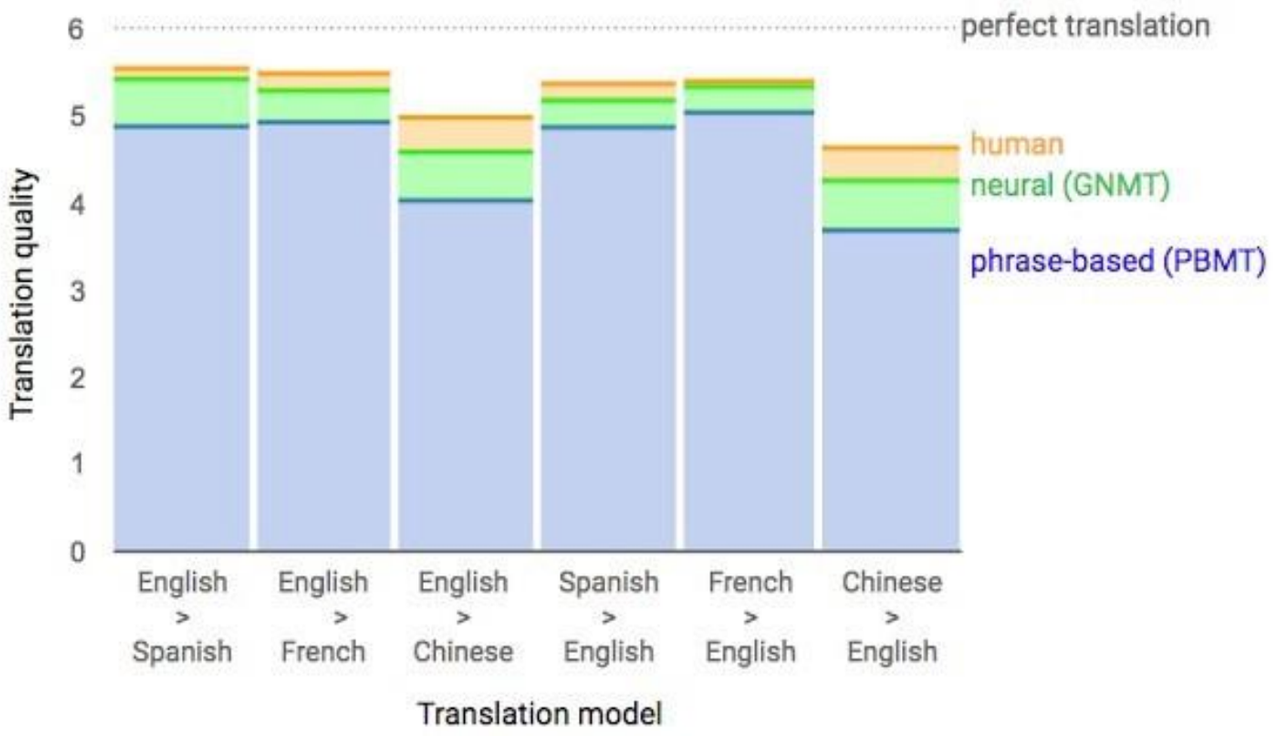
Изображение в изображение Pix2pix



Zhu и др. «Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks» 2018 // https://arxiv.org/pdf/1703.10593.pdf

Обработка текста

- **Понимание смысла (Language Understanding)**
- **Автоматический перевод**

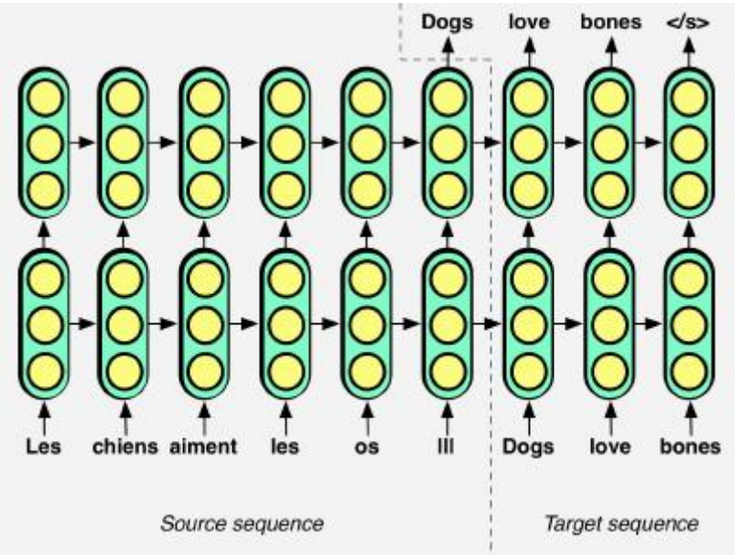


Google's Neural Machine Translation System

https://arxiv.org/abs/1609.08144

Обработка текста

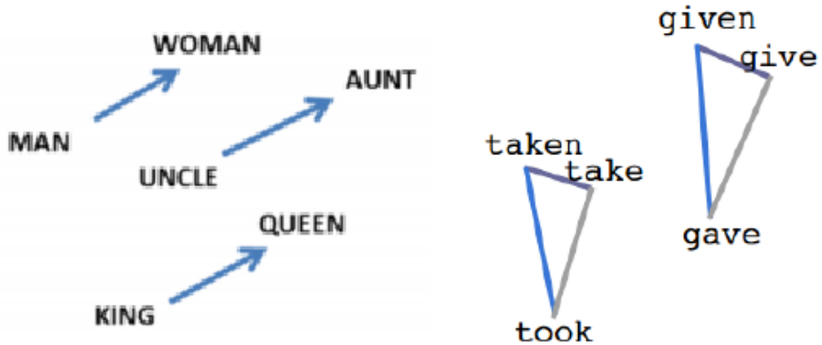
- **Диалоговые системы**



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

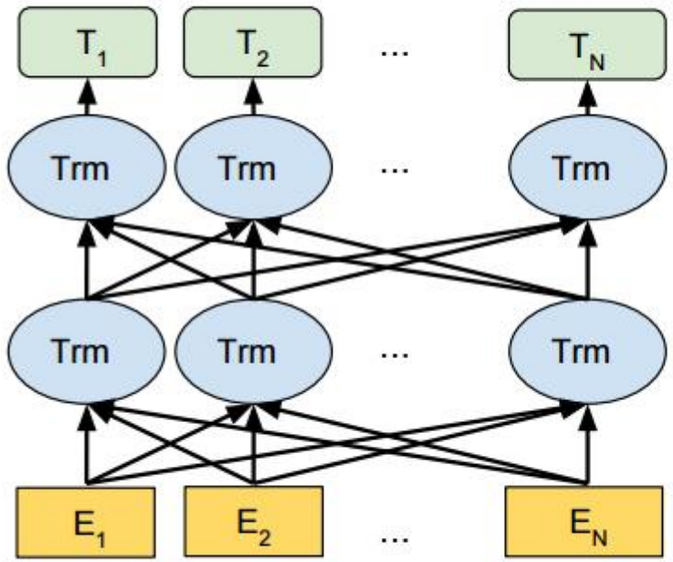
Обработка текста

Векторизация слов: word2vec, GloVe, ...



Woman – Man ≈ Aunt - *Uncle*
King - Male + Female ≈ *Queen*
Human - Animal ≈ *Ethics*

Обработка текста



новые архитектуры

(GPT BERT XLNET Roberta ALBERT)

Языковые модели / генерация текста

Порфирьевич

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/

Генерация текста

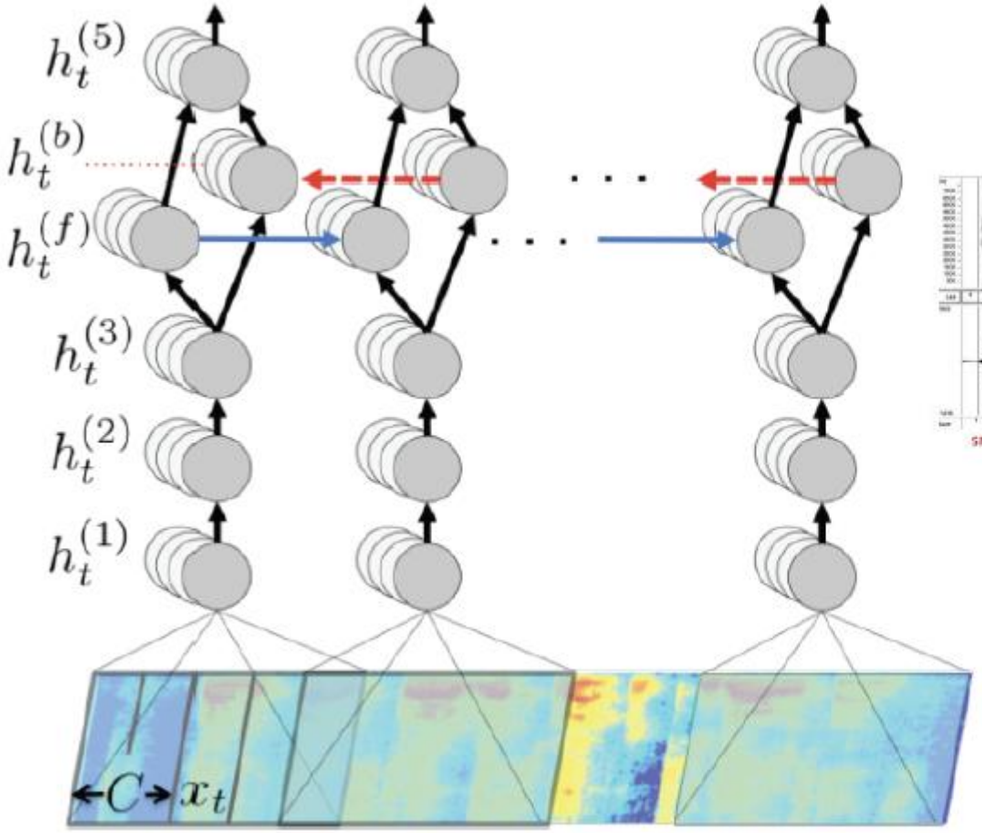
This is an impressive demo of a recurrent neural network.

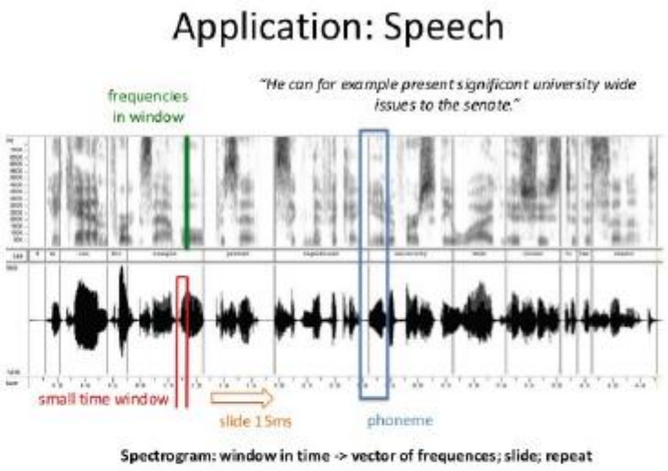
This is an impressive demo of a recurrent neural network.

This is an impressive demo of a recurrent neural network

Обработка аудио / звуков / речи

- **Распознавание речи (Speech Recognition) – уже было**
-





Обработке 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

DataSet: http://www.robots.ox.ac.uk/~vgg/data/lip_reading/lrs2.html

Обработке видео / аудио

Deepfake — методика синтеза изображения, основанная на искусственном интеллекте

Обработке видео / аудио

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

Обработка видео / 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
Google

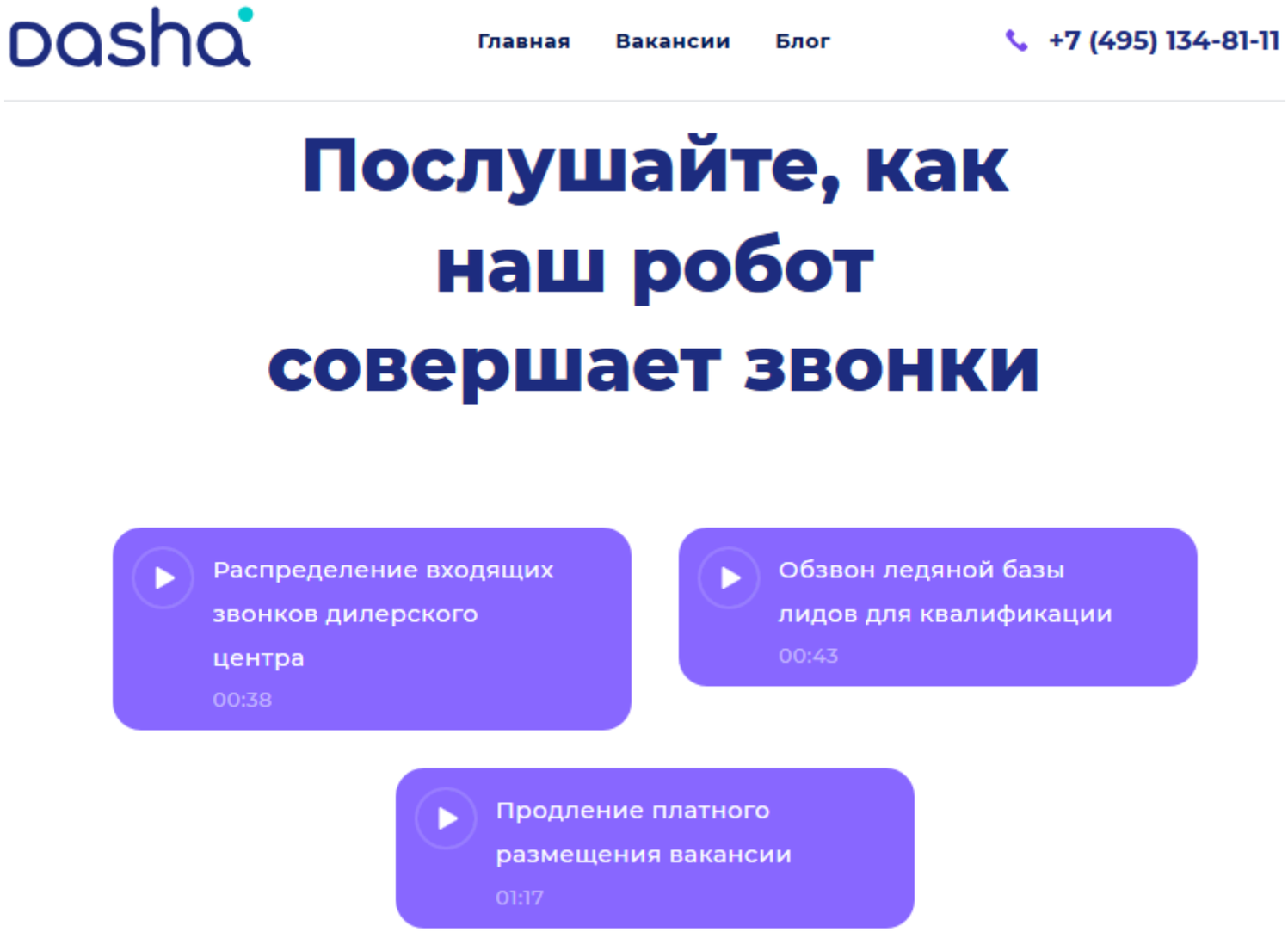
- произношение
- громкость
- скорость
- интонация

Mellotron (based on Tacotron 2 GST)
NVIDIA

- трансфер ритма
- голос из аудио

https://research.googleblog.com/2017/12/tacotron-2-generating-human-like-speech.html
https://nv-adlr.github.io/Mellotron

Автоматизация колл-центров / разговорный ИИ



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: Pix2Code

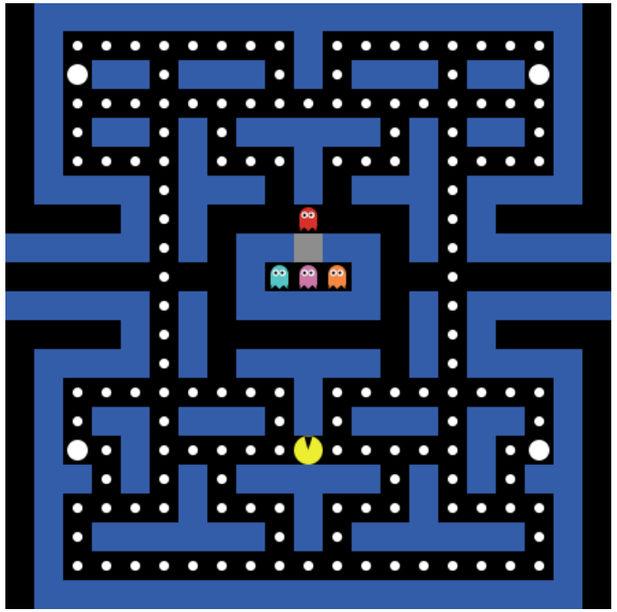
(a) iOS GUI screenshot

```
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.

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

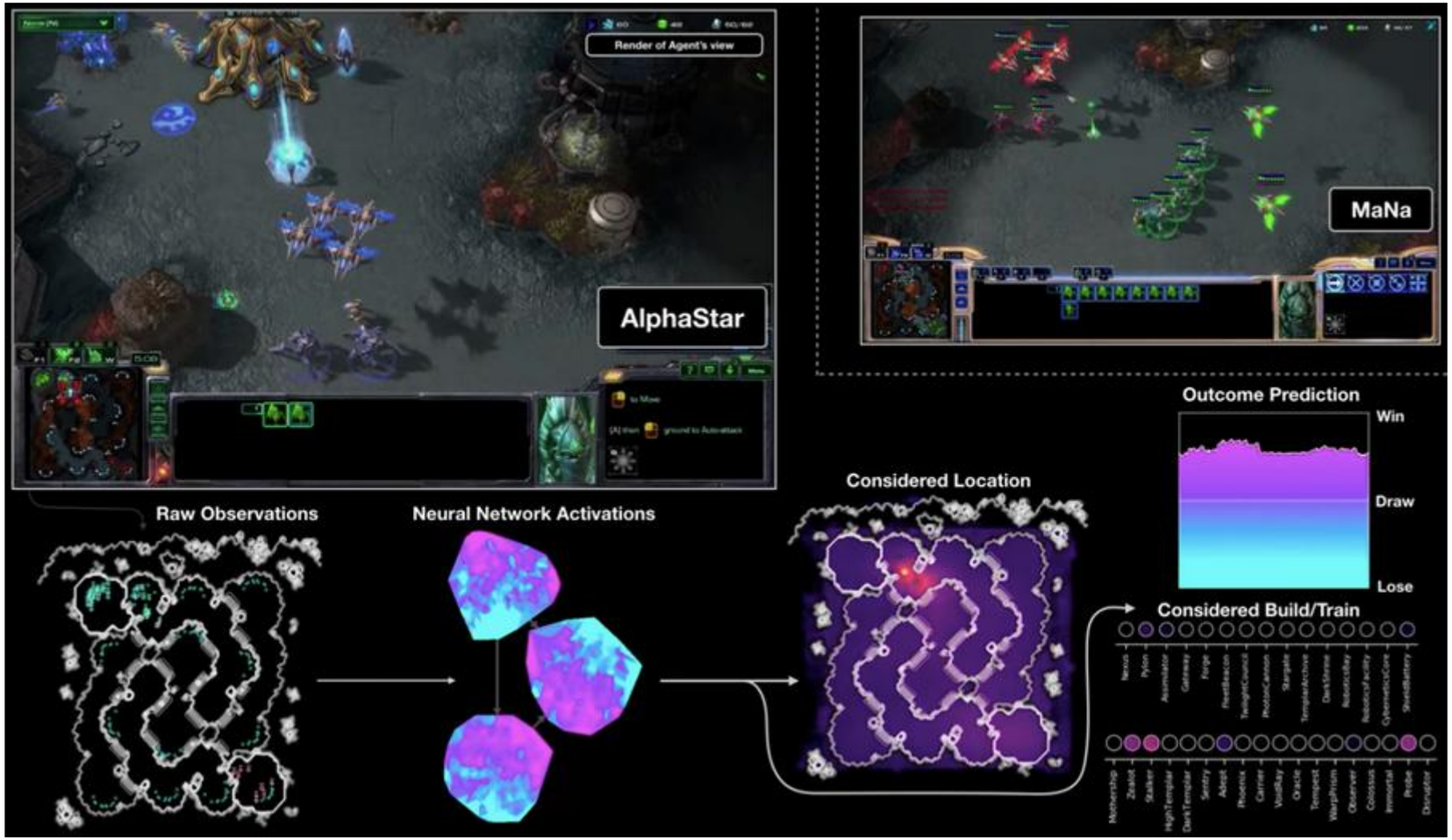
1920 процессоров и 280 графических процессоров

ИИ в играх

Alpha Go Zero: Играет без анализа человеческого опыта

https://arxiv.org/abs/1712.01815

AlphaStar от DeepMind для StarCraft II

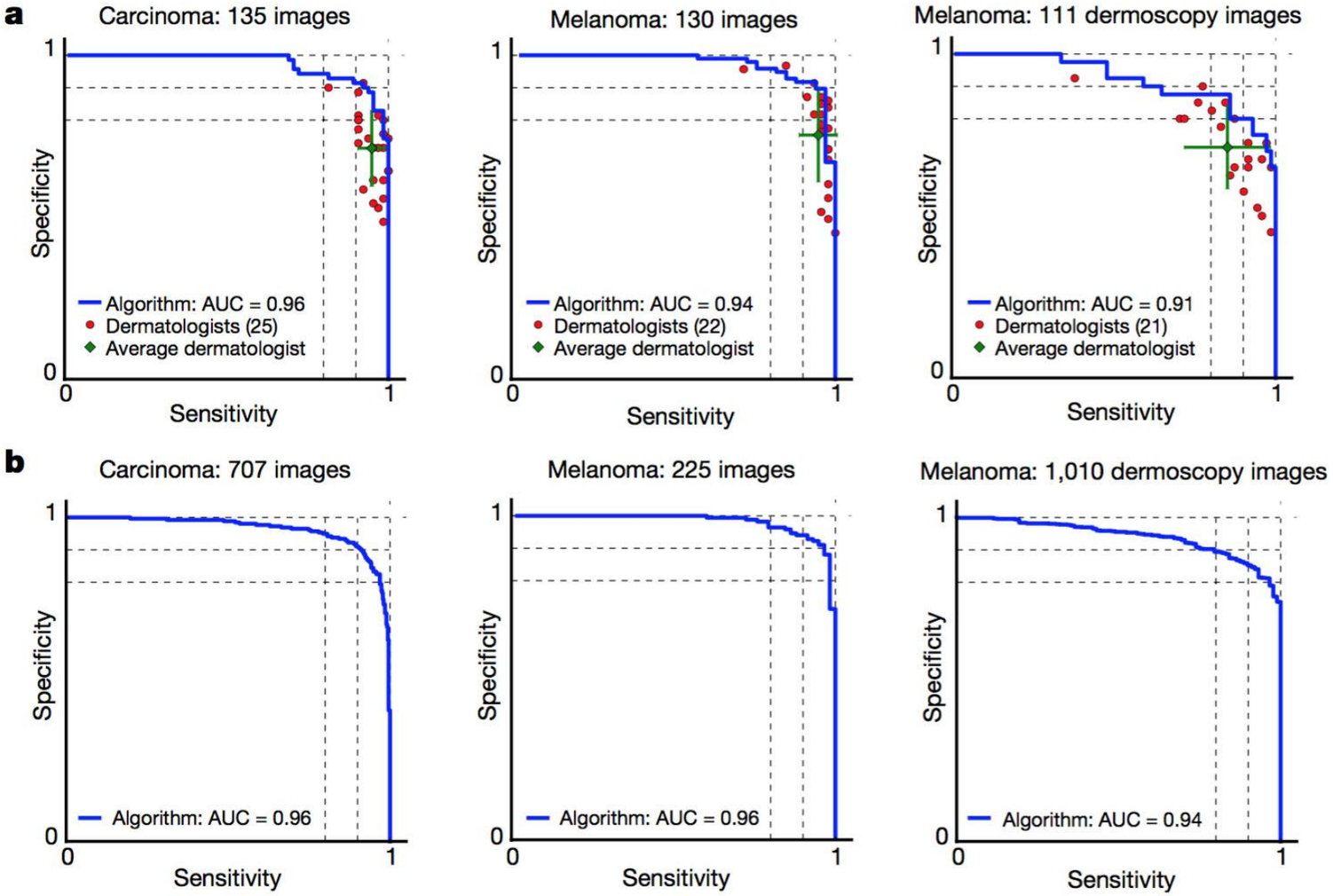


19/12/12 – AlphaStar – Grzegorz Komincz (MaNa) из команды Liquid – 5:0

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

Медицина (Drug Discovery and Medical)

Дерматология: диагностика рака кожи



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

Медицина (Drug Discovery and Medical)

CNN на низкоразмерной геометрической структуре

AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery

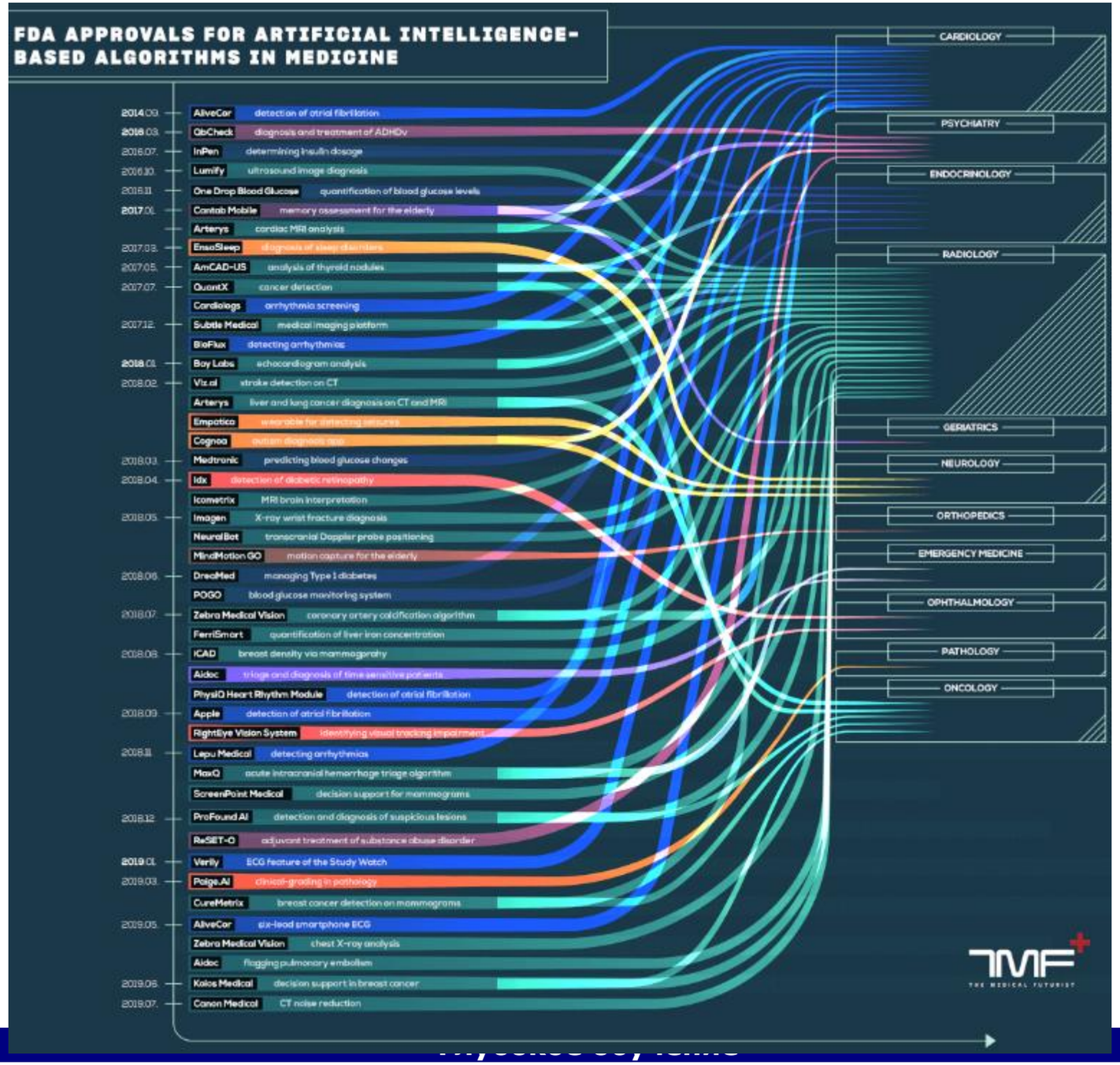
Izhar Wallach
Atomwise, Inc.
izhar@atomwise.com

Michael Dzamba
Atomwise, Inc.
misko@atomwise.com

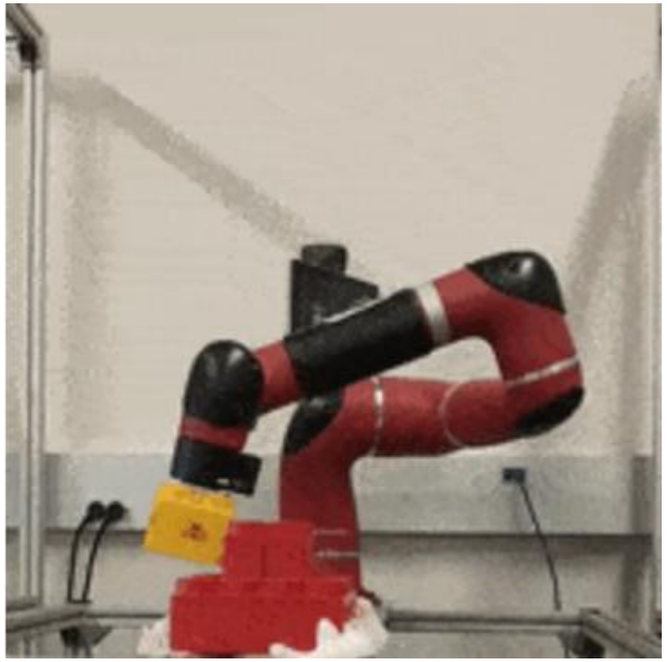
Abraham Heifets
Atomwise, Inc.
abe@atomwise.com

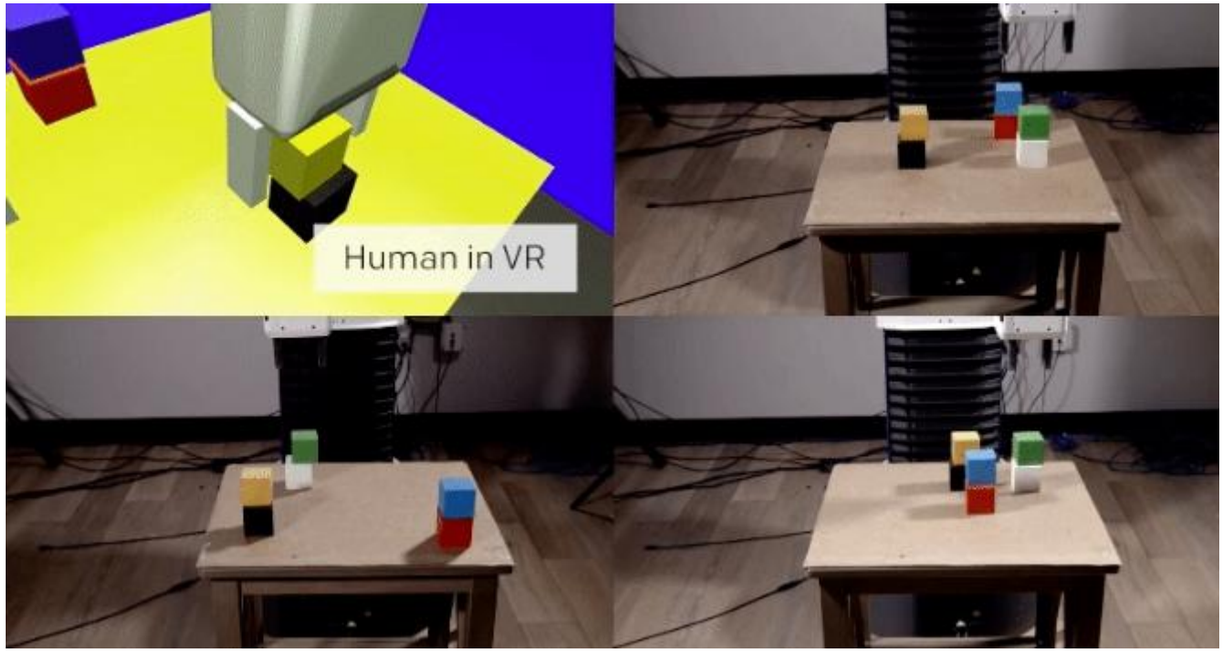
Abstract

Deep convolutional neural networks comprise a subclass of deep neural networks (DNN) with a constrained architecture that leverages the spatial and temporal structure of the domain they model. Convolutional networks achieve the best predictive performance in areas such as speech and image recognition by hierarchically composing simple local features into complex models. Although DNNs have


FDA APPROVALS FOR ARTIFICIAL INTELLIGENCE-BASED ALGORITHMS IN MEDICINE
2014.09. AliveCor detection of atrial fibrillation
2016.03. QbCheck diagnosis and treatment of ADHDv
2016.07. InPen determining insulin dosage
2016.10. Lumify ultrasound image diagnosis
2016.11. One Drop Blood Glucose quantification of blood glucose levels
2017.01. Cantab Mobile memory assessment for the elderly
Arterys cardiac MRI analysis
2017.03. EnseSleep diagnosis of sleep disorders
2017.05. AmCAD-US analysis of thyroid nodules
2017.07. QuantX cancer detection
Cardiologs arrhythmia screening
2017.12. Subtle Medical medical imaging platform
BioFlux detecting arrhythmias
2018.01. Bay Labs echocardiogram analysis
2018.02. Viz.ai stroke detection on CT
Arterys liver and lung cancer diagnosis on CT and MRI
Empatica wearable for detecting seizures
Cognoa autism diagnosis app
2018.03. Medtronic predicting blood glucose changes
2018.04. Idx detection of diabetic retinopathy
Icometrix MRI brain interpretation
2018.05. Imagen X-ray wrist fracture diagnosis
NeuralBot transcranial Doppler probe positioning
MindMotion GO motion capture for the elderly
2018.06. DreaMed managing Type 1 diabetes
POGO blood glucose monitoring system
2018.07. Zebra Medical Vision coronary artery calcification algorithm
FerriSmart quantification of liver iron concentration
2018.08. ICAD breast density via mammography
Aidoc triage and diagnosis of time-sensitive patients
PhysiQ Heart Rhythm Module detection of atrial fibrillation
2018.09. Apple detection of atrial fibrillation
RightEye Vision System identifying visual tracking impairment
2018.11. Lepu Medical detecting arrhythmias
MaxQ acute intracranial hemorrhage triage algorithm
ScreenPoint Medical decision support for mammograms
2018.12. ProFound AI detection and diagnosis of suspicious lesions
ReSET-O adjuvant treatment of substance abuse disorder
2019.01. Verily ECG feature of the Study Watch
2019.03. Paige.AI clinical-grading in pathology
CureMetrix breast cancer detection on mammograms
2019.05. AliveCor six-lead smartphone ECG
Zebra Medical Vision chest X-ray analysis
Aidoc flagging pulmonary embolism
2019.06. Koios Medical decision support in breast cancer
2019.07. Canon Medical CT noise reduction
CARDIOLOGY
PSYCHIATRY
ENDOCRINOLOGY
RADIOLOGY
GERIATRICS
NEUROLOGY
ORTHOPEDICS
EMERGENCY MEDICINE
OPHTHALMOLOGY
PATHOLOGY
ONCOLOGY
TMF
THE MEDICAL FUTURIST

Автомобили-беспилотники, роботы-манипуляторы

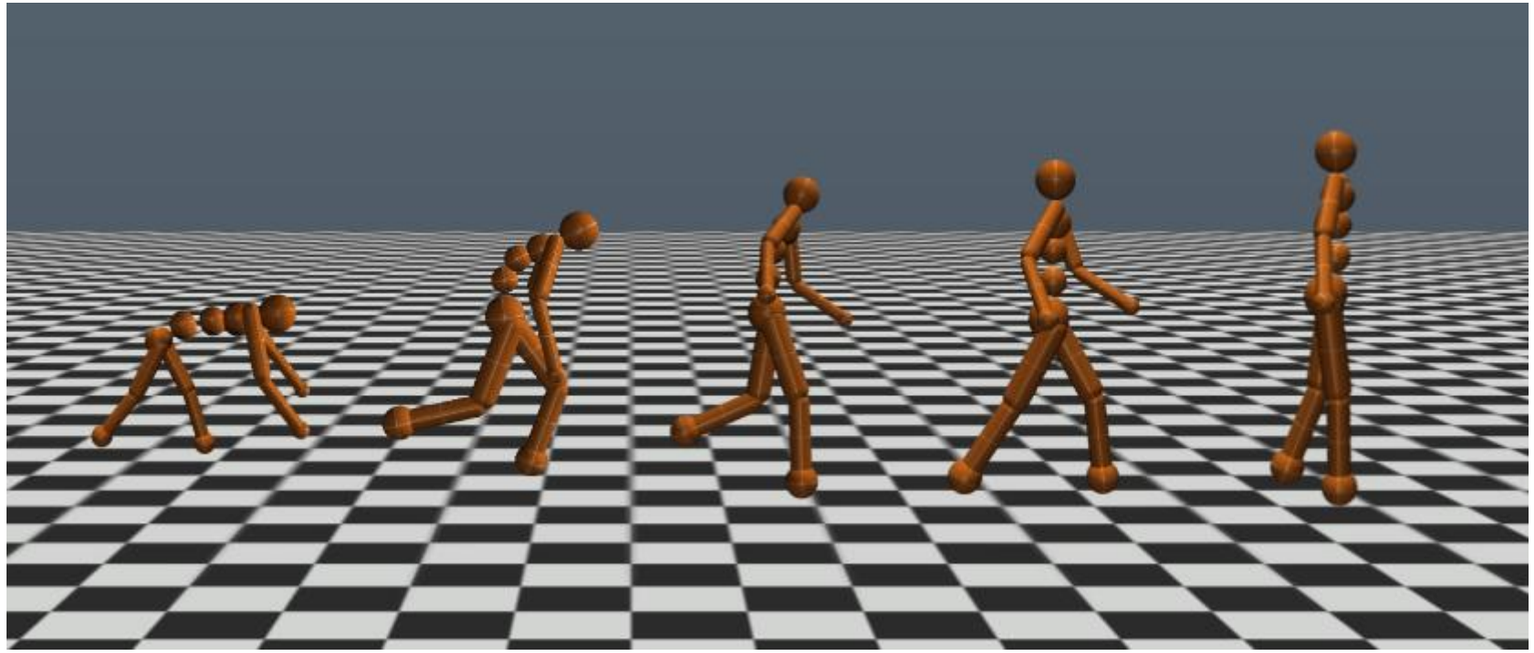




http://bair.berkeley.edu/blog/2017/10/06/soft-q-learning/

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

Нейроэволюционные алгоритмы



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

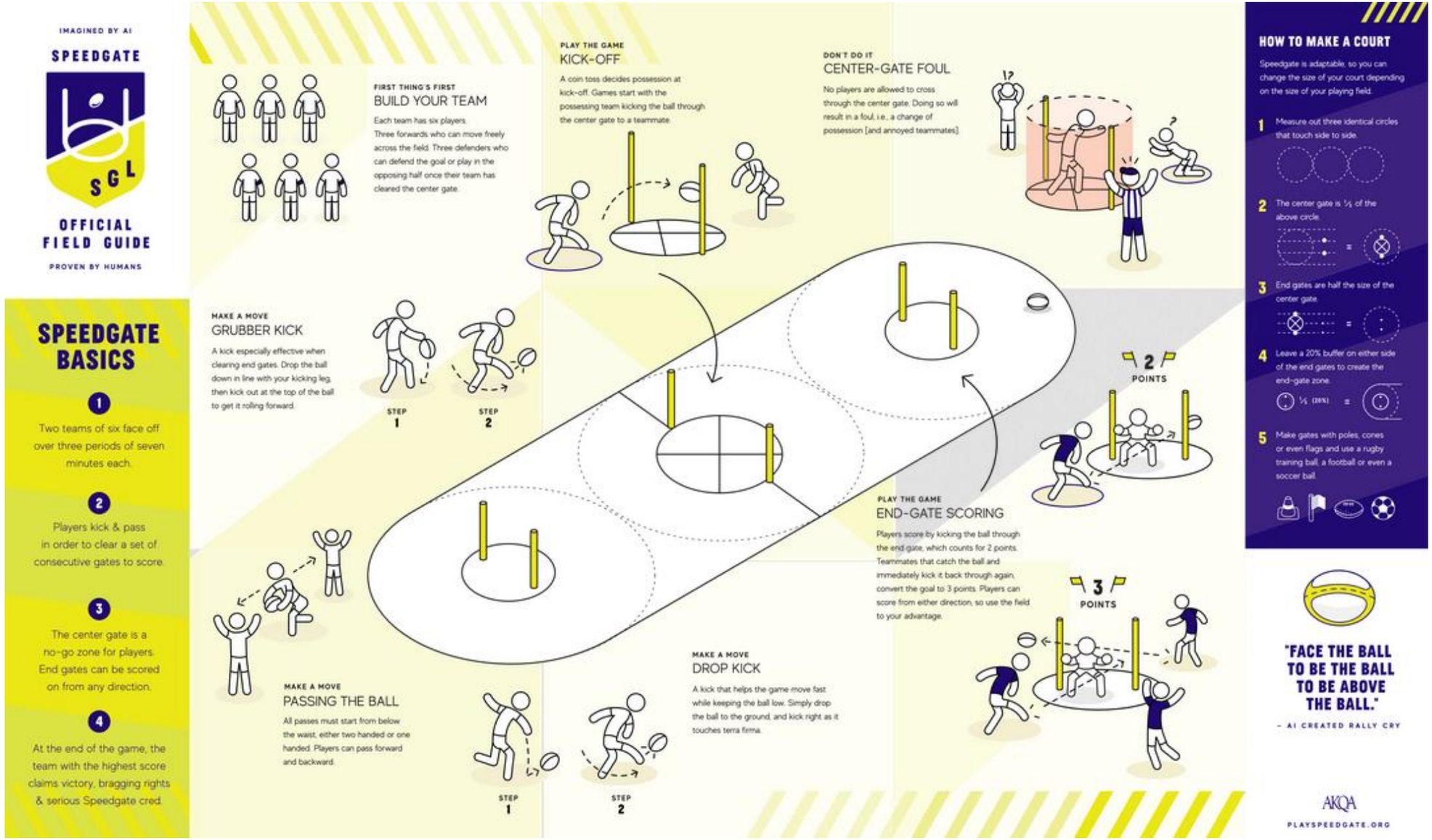
Подбор гардероба

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

Реконструкция изображений / панорам

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

Изобретение нового вида спорта



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

AutoML: пусть DL делает всё!

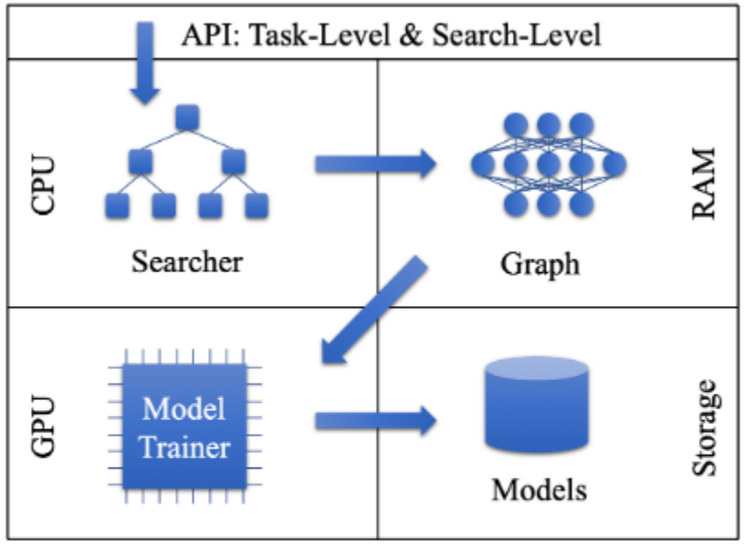


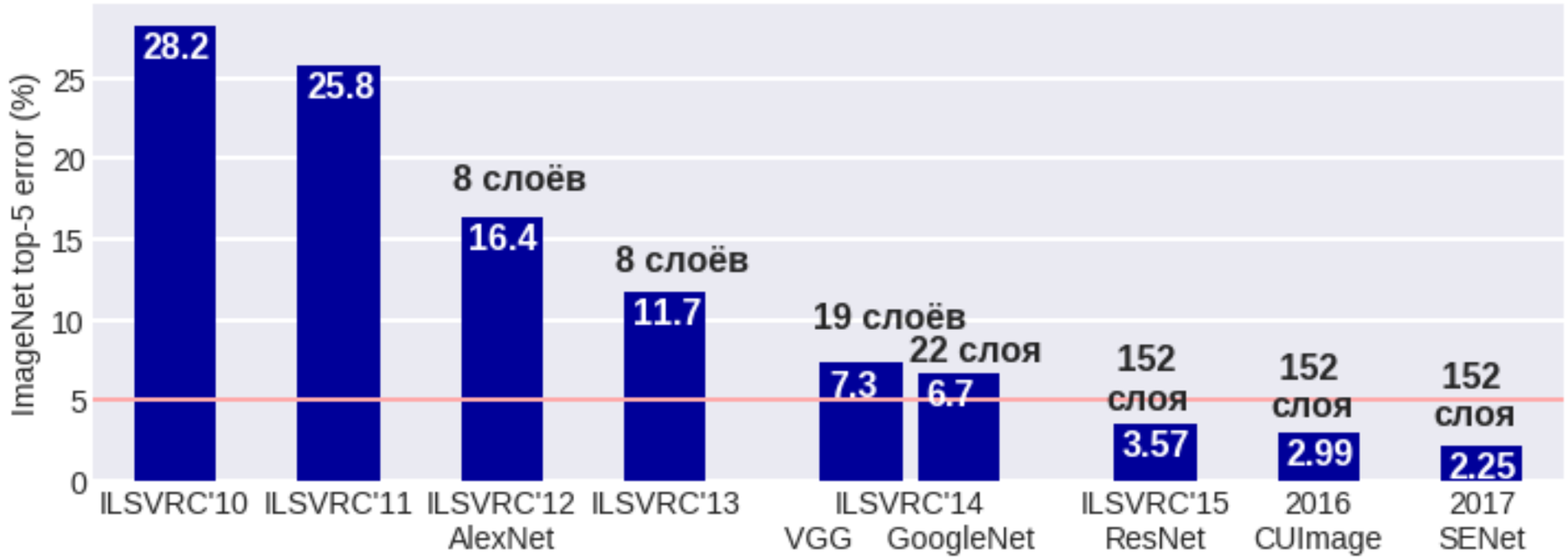
Figure 2: Auto-Keras System Overview. (1) User calls the API. (2) The Searcher generates neural architectures on CPU. (3) Graph builds real neural networks with parameters on RAM from the neural architectures. (4) The neural network is copied to GPU for training. (5) Trained neural networks are saved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

Революция в машинном обучении



ошибка человека – 5.1

Причины прогресса

- **Интенсивные исследования в ML (+ обмен идеями)**
- **Огромные наборы данных (+ интеренет, 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 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	2x2550 = 5100	2x637 = 1274

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

Проблемы 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**

Минусы нейросетей

- Большие выборки
- Не учитывают «естественный контекст» (пока не ИИ) / не учат инварианты
- Пока проблемы с интерпретаций (почему такой ответ)
- Уязвимости (возможность атак)
- Ресурсоёмкость
- Проблемы с выбором архитектуры

сеть	ошибка	время
ResNet18	0.108	2.5 дня
ResNet50	0.07	5 дней
ResNet101	0.062	1 неделя

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

История

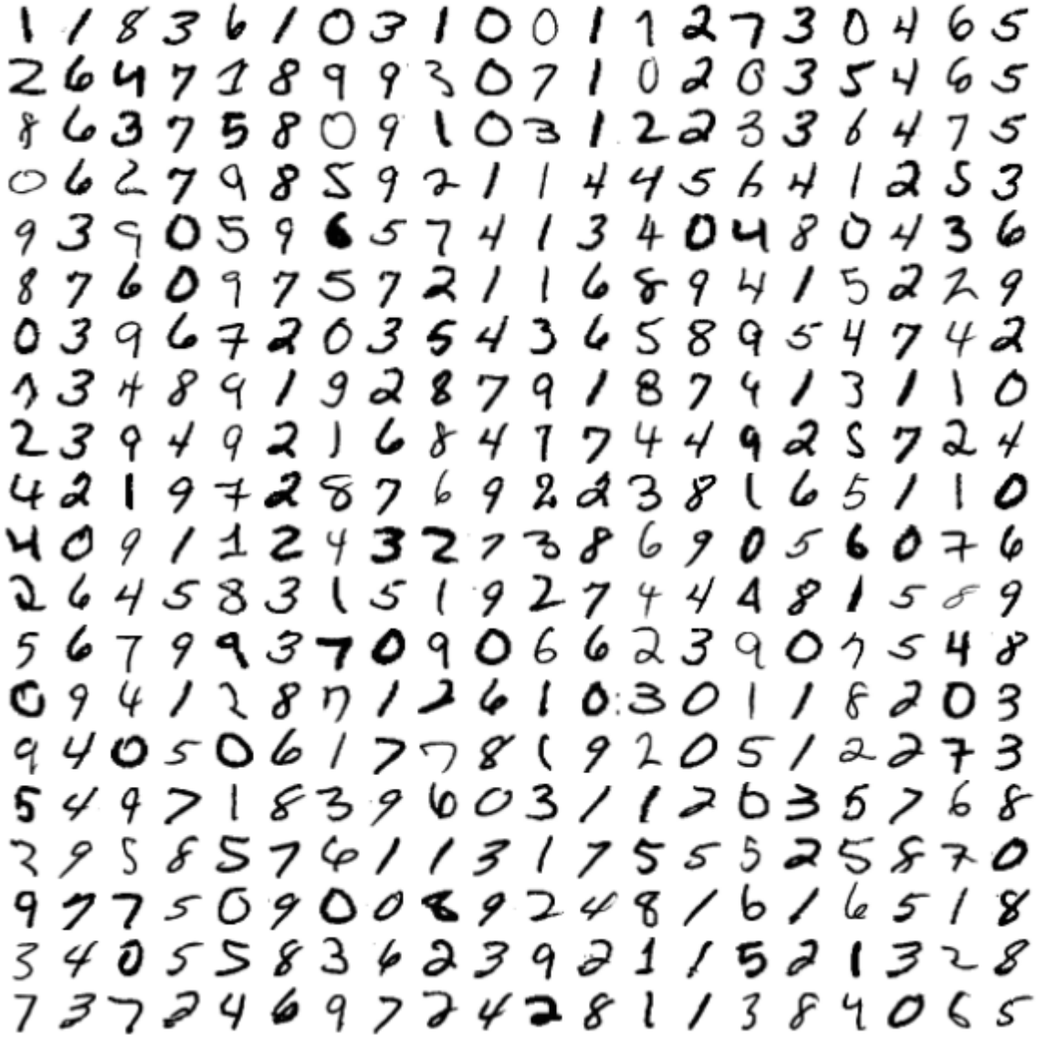
1950 – 70 – «коннективизм», модель нейрона
1970 – 80 – зима ИИ
1980 – 1990 – экспертные системы, обратное распространение
1990 – 2000 – зима
1997 – LSTM (Hochreiter, Schmidhuber)
1998 – распознавание символов ConvNets (LeCun)
2006 – Deep Boltzmann Machines (Hinton, Salakhutdinov)
2009 – распознавание речи
2012 – Победа на ImageNet ConvNets (Krizhevsky и др.)
2014 – информационный / инвестиционный бум

ПО

	Language(s)	License	Main backer
PyTorch	**Python**	BSD	Facebook
Caffe2	C++, Python	Apache	Facebook
TensorFlow	Python, C++	Apache	Google
MXNet	Python, C++, R, Scala	Apache	Amazon
CNTK	Python, C++	MIT	Microsoft
Torch	Lua	BSD	Facebook
Theano	Python	BSD	U. of Montreal
Caffe	C++	BSD 2 clauses	U. of CA, Berkeley

MNIST

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



Fashion-MNIST

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

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

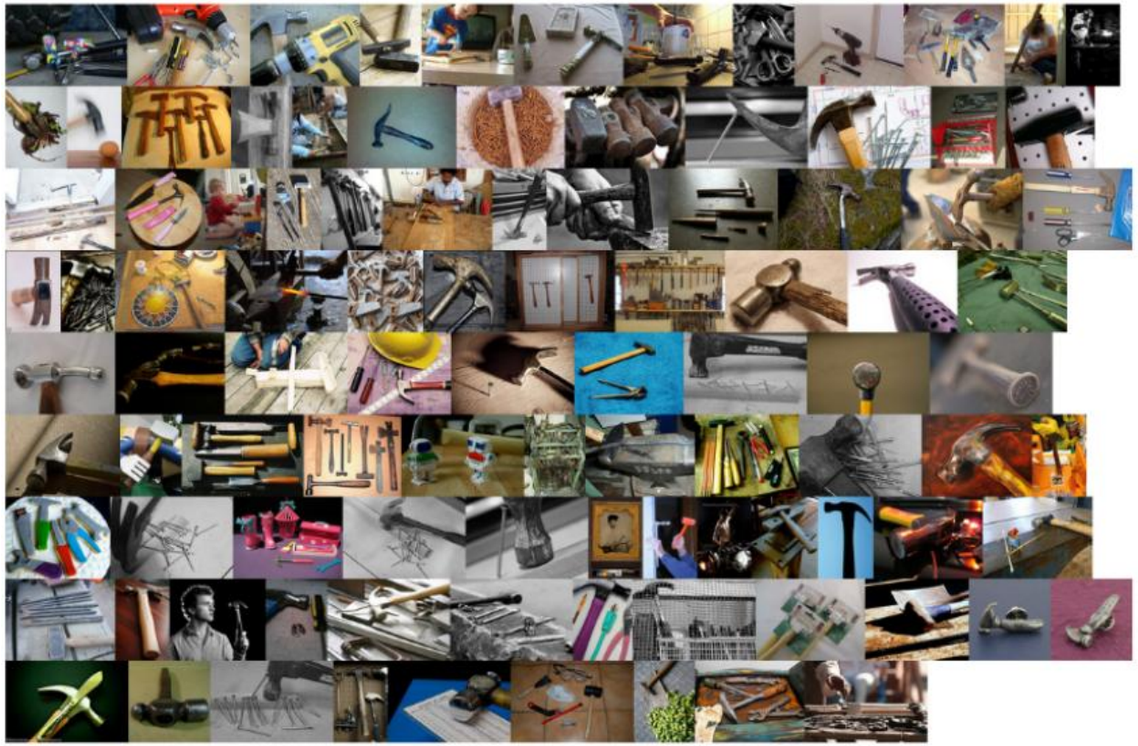
ImageNet

IMAGENET

1.2 млн. 225x 225-изображений, 1000 классов

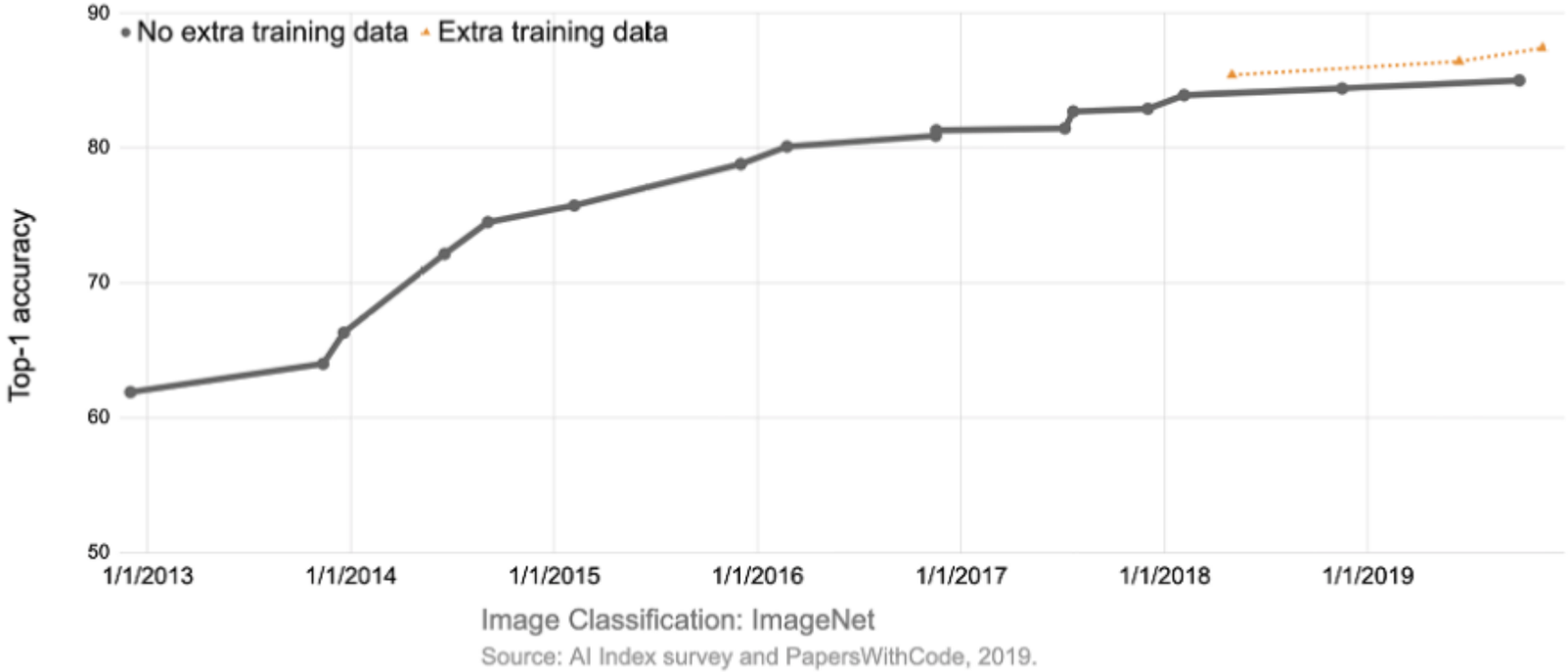
есть 12 млн. 21000 классов

Класс «Молотки»



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

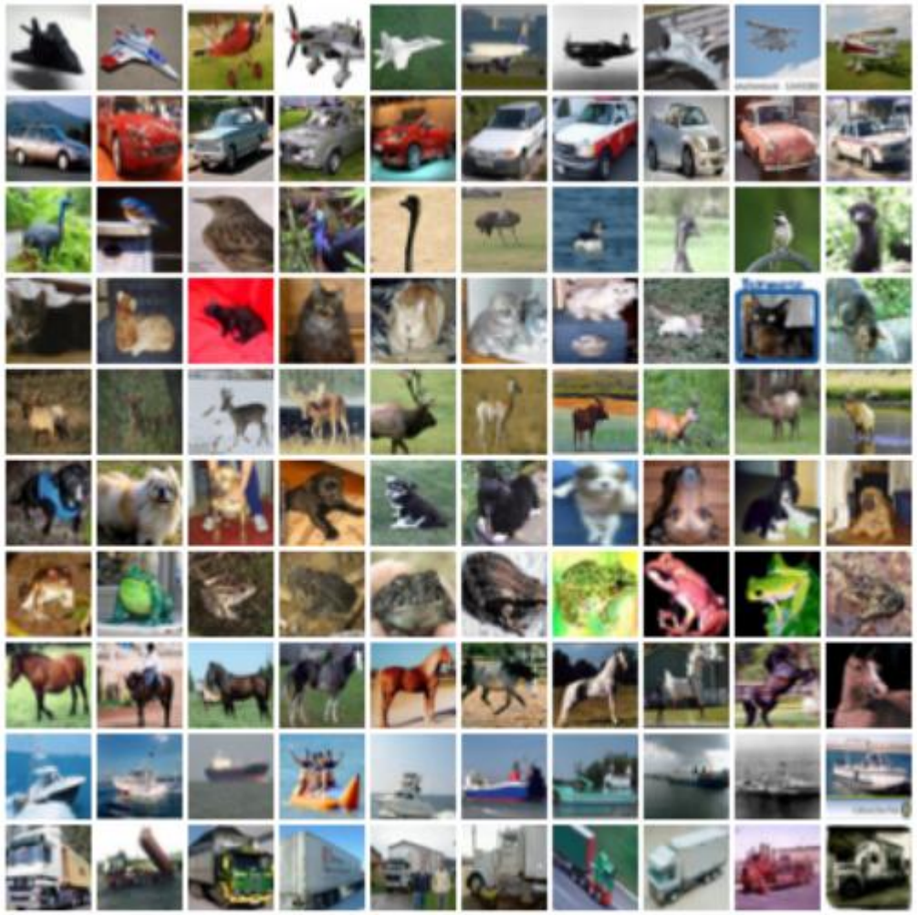
ImageNet

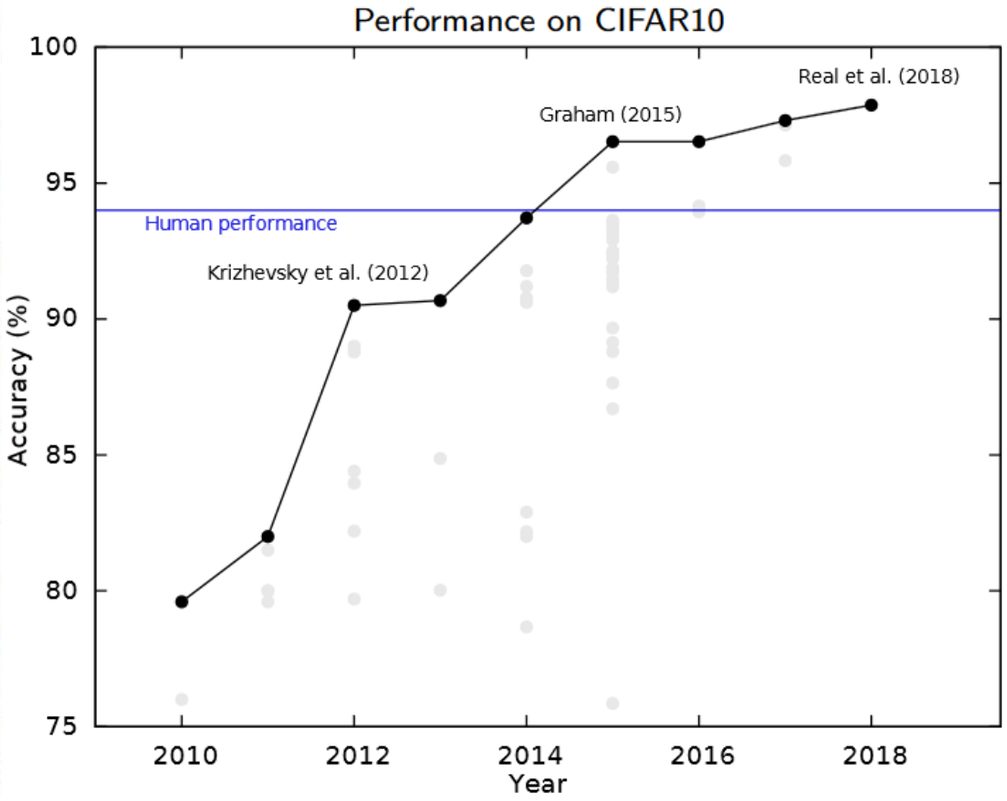


[Artificial Intelligence Index Report 2019]

CIFAR10

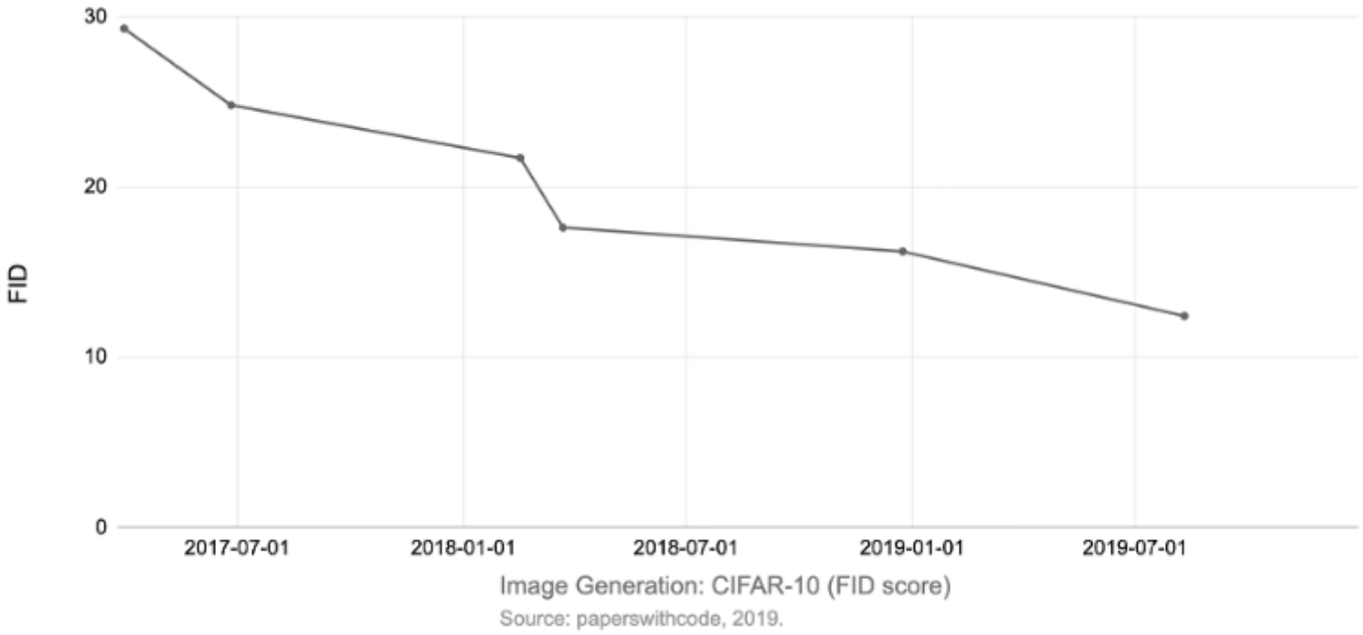
32×32 цветные, 50k – обучение, 10k – тест





10 классов: самолет, автомобиль, птица, кошка, олень, собака, лягушка, лошадь, корабль и грузовик

CIFAR10



Генерация изображений. Метрика – расстояние в признаковом пространстве Inception v3

Places

places

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

Microsoft COCO

COCO
Common Objects in Context

- 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

http://cocodataset.org/#home

Изображения – датасеты

PASCAL VOC

CIFAR-10

CIFAR-100

ImageNet

Fashion MNIST

MS COCO

VoxCeleb

Open Images

CINIC-10

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

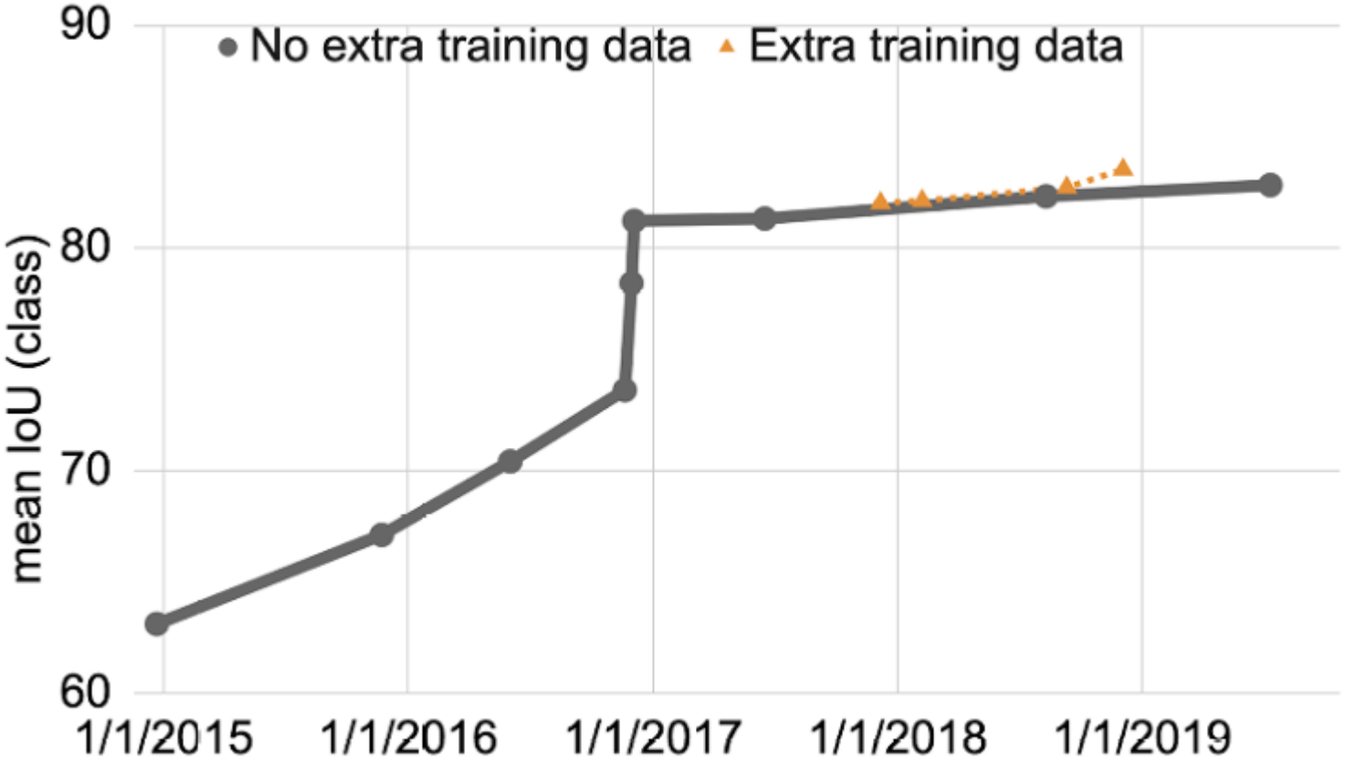
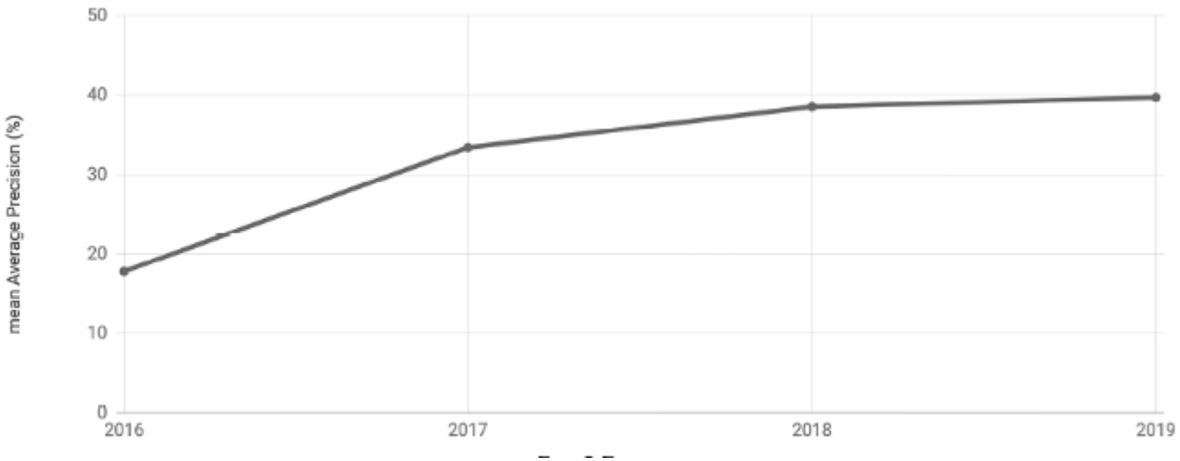


Fig. 3.4.
Note: The orange dots denote tests with additional training data.

семантическая сегментация

ActivityNet



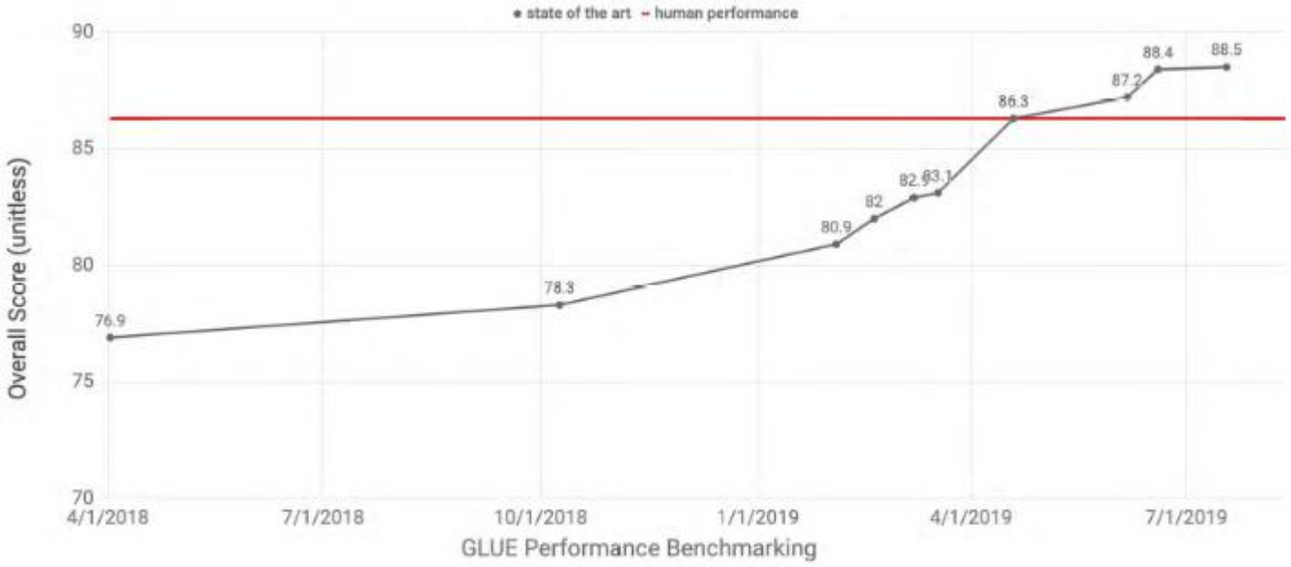
A Large-Scale Video Benchmark for Human Activity Understanding

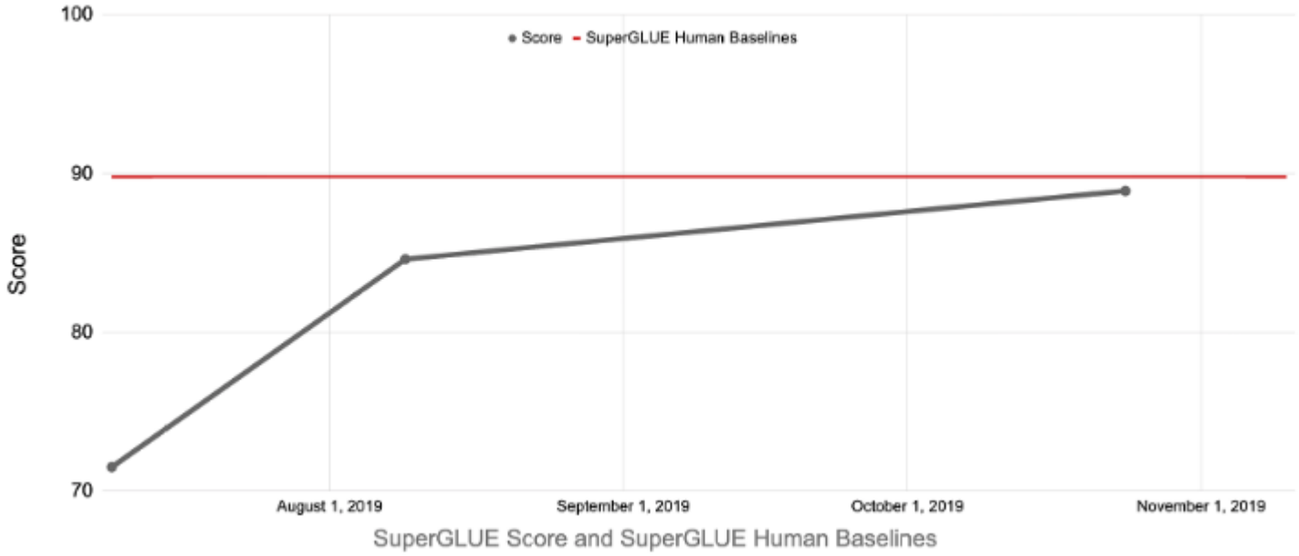
http://activity-net.org

200 activity classes

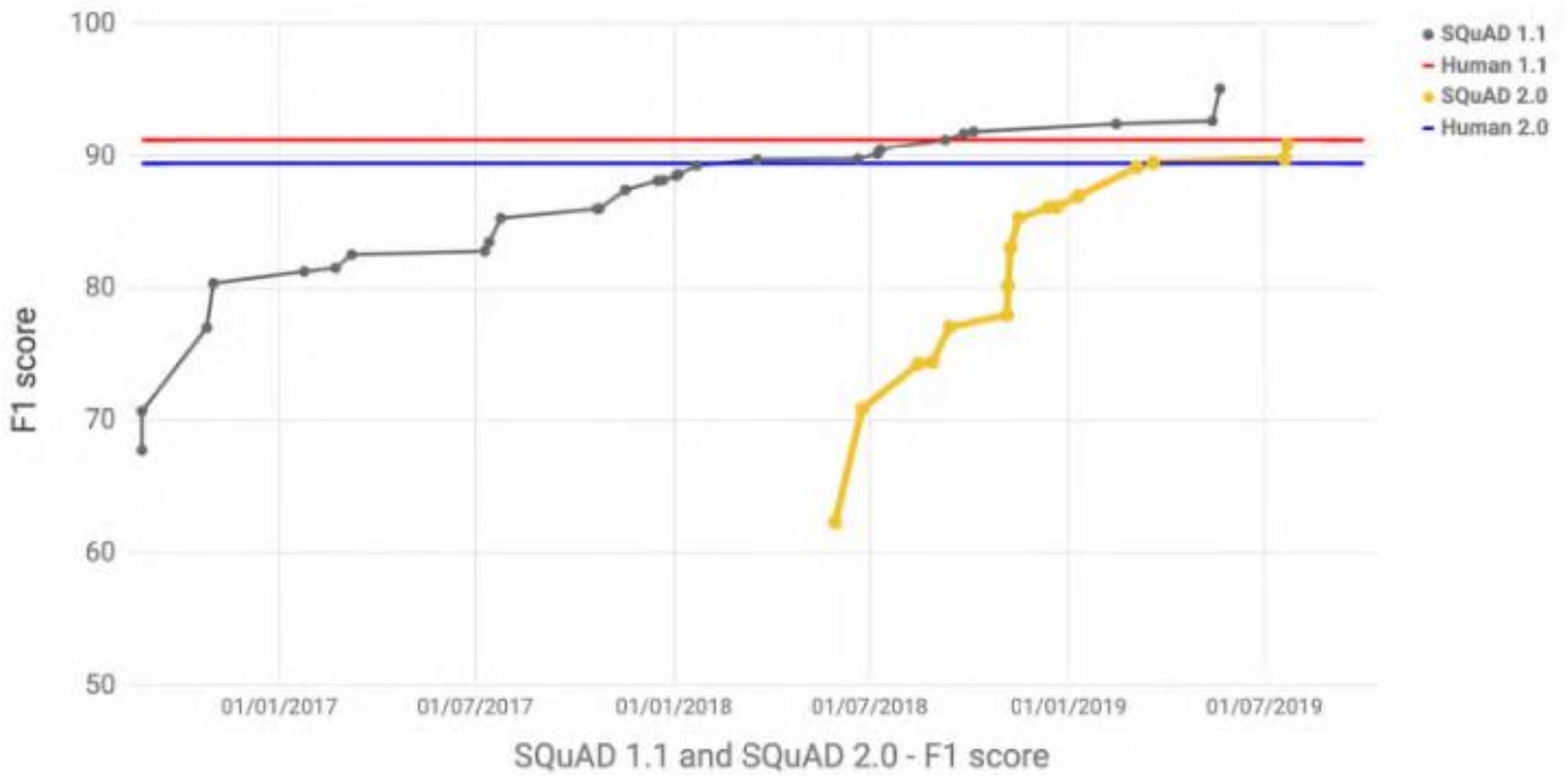
- **10,024 training videos (15,410 instances)**
- **4,926 validation videos (7,654 instances)**
- **5,044 testing videos (labels withheld)**

GLUE (General Language Understanding Evaluation) / superGLUE





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

Глубокое обучение (Deep learning) – обучение многослойных моделей:

- сетей прямого распространения
- многослойных графических моделей

(deep belief network, deep Boltzmann machine)

	генерация признаков → модель
Классическое МО	генерация признаков → нелинейности (kernel), преобразования (PCA), селекция → модель
Глубокое обучение	модель

Пример: MNIST на Pytorch

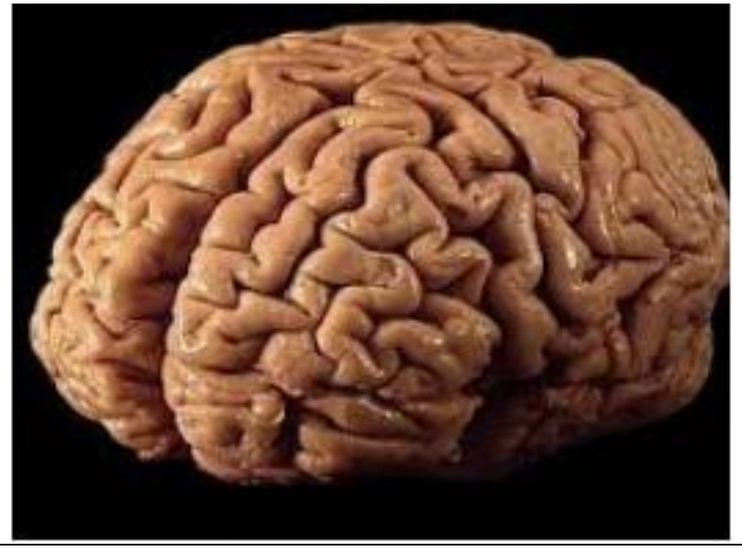
```
model = nn.Sequential(
    nn.Conv2d( 1, 32, 5), nn.MaxPool2d(3), nn.ReLU(),
    nn.Conv2d(32, 64, 5), nn.MaxPool2d(2), nn.ReLU(),
    Flattener(),
    nn.Linear(256, 200), nn.ReLU(),
    nn.Linear(200, 10)
)

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

AI vs человек: когда компьютер превзошёл человека…

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

Мозг vs процессор

	Мозг	Процессор
		
Производительность, бит/с	**~10^{20}**	**~10^{19}**
энергопотребление	**~ 20 Вт**	**~ 200 Вт**

Пока процессоры проигрывают, но уже близко…

Другие курсы

Классика DL CS231n:
«Convolutional Neural Networks for Visual Recognition from Stanford»
http://cs231n.stanford.edu/syllabus.html

Hugo Larochelle Neural Networks // Deep Learning Summer School, Montreal 2016
http://videolectures.net/deeplearning2016_larochelle_neural_networks/

Антон Осокин Материалы курса «Глубинное обучение»// ФКН ВШЭ, 2018
https://github.com/aosokin/DL_CSHSE_spring2018

Лучшая книга по глубокому обучению

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

Документация

https://www.tensorflow.org/tutorials/
https://pytorch.org/tutorials/

Датасеты

https://www.datasetlist.com/

Текущие рекорды (но уже не обновляется)

https://paperswithcode.com
https://www.stateoftheart.ai
https://sotabench.com
https://www.eff.org/ai/metrics

Поиграться с NN

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