САМАРСКИЙ УНИВЕРСИТЕТ
SAMARA UNIVERSITY

Нейронные сети и глубокое обучение: вводная часть

Артем Владимирович Никоноров

д.т.н., руководитель лаборатории интеллектуального анализа видеоданных,
Института систем обработки изображений РАН,
директор Института искусственного интеллекта

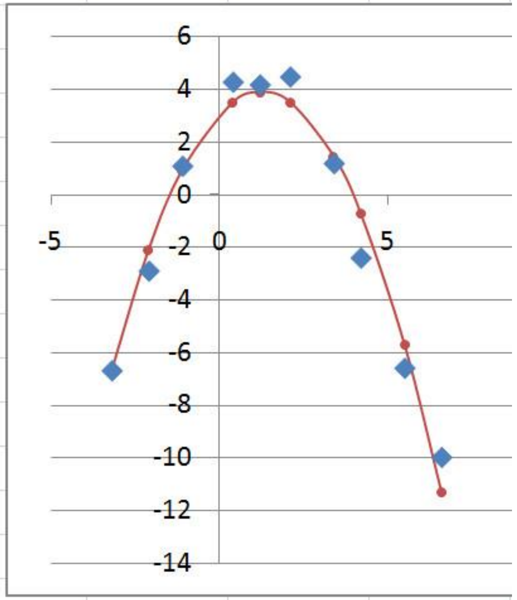
Самарского Университета

Самара
2023

Простейший пример: подбор параметров модели

Задача – аппроксимировать точки кривой

i	xi	yi	y=a0+a1*x+a2*x^2	Error=yi-y
1	-3.2	-6.7	-6.7212	0.021
2	-2.1	-2.9	-2.1288	-0.771
3	-1.1	1.1	0.9249	0.175
4	0.4	4.3	3.5033	0.797
5	1.2	4.2	3.8961	0.304
6	2.1	4.5	3.5210	0.979
7	3.4	1.2	1.4523	-0.252
8	4.2	-2.4	-0.7177	-1.682
9	5.5	-6.6	-5.7016	-0.898
10	6.6	-10	-11.3282	1.328
Standard deviation				**0.9295417**



$$f(\mathbf{a}, \mathbf{x}_{i,}) = \mathbf{y}_i, \; i = 1..N$$

$$a_2 \mathbf{x}_i^2 + a_1 \mathbf{x}_1 + a_0 = \mathbf{y}_i + e_i$$

$$\mathbf{a} = \arg\min_{\mathbf{a}} \sum_{i=1}^{N} \left(a_2 \mathbf{x}^2 + a_1 \mathbf{x} + a_0 - \mathbf{y}_i \right)^2$$

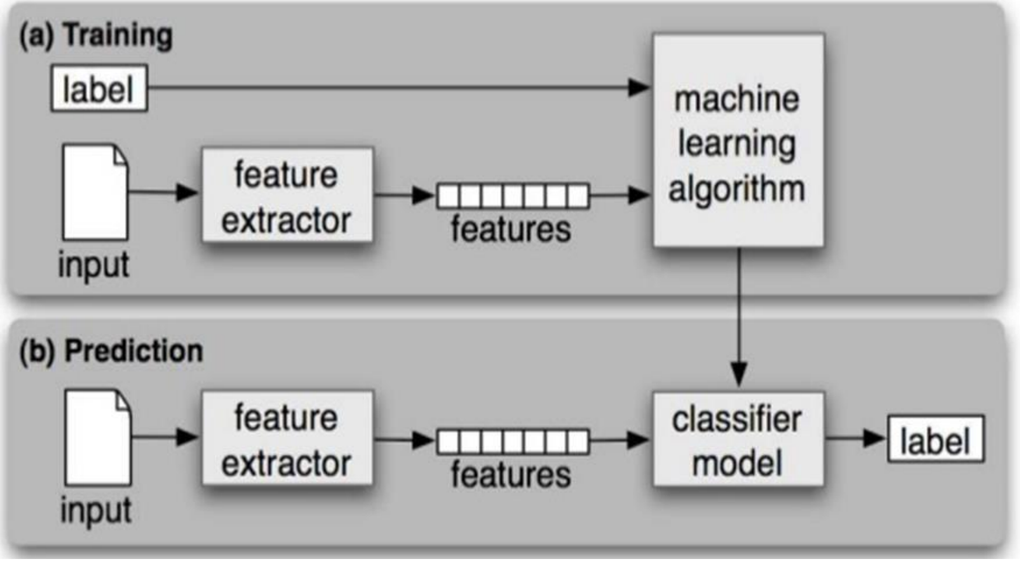
MACHINE LEARNING – Data Driven Approach

Обучение – нахождение зависимостей в данных,
Цель - построение прогноза по имеющимся данным

Supervised Learning / Обучение с учителем

$$f(\mathbf{a}, \mathbf{x}_{i,}) = \mathbf{y}_i, \ i = 1..N$$ Обучающая выборка

$$f(\mathbf{a}, \mathbf{x}_{N+1}) = ?$$ Прогноз (инференс)

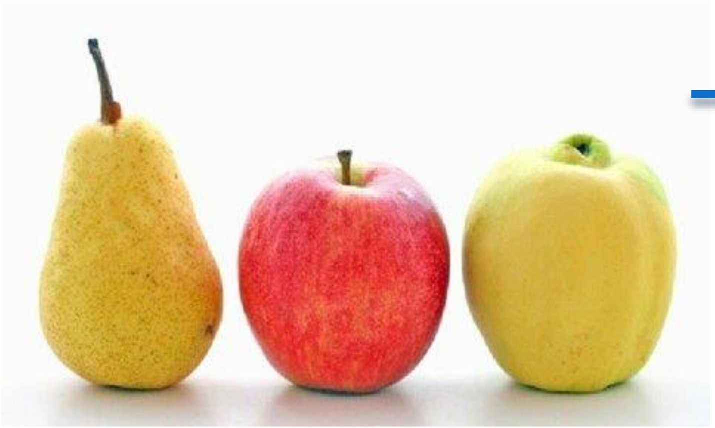


Варианты:
Unsupervised, Semi-supervised

Y – вещественное – регрессия
натуральное – классификация
0,1 – двухклассовая классификация

Классификация, или распознавание основано на признаках (features)

Поиск признаков – искусство, эвристика, инженерная интуиция
Классификация – математика
Становление одного из подходов от Ю.И. Журавлева:
https://www.youtube.com/watch?v=R3CMqrrIWOk

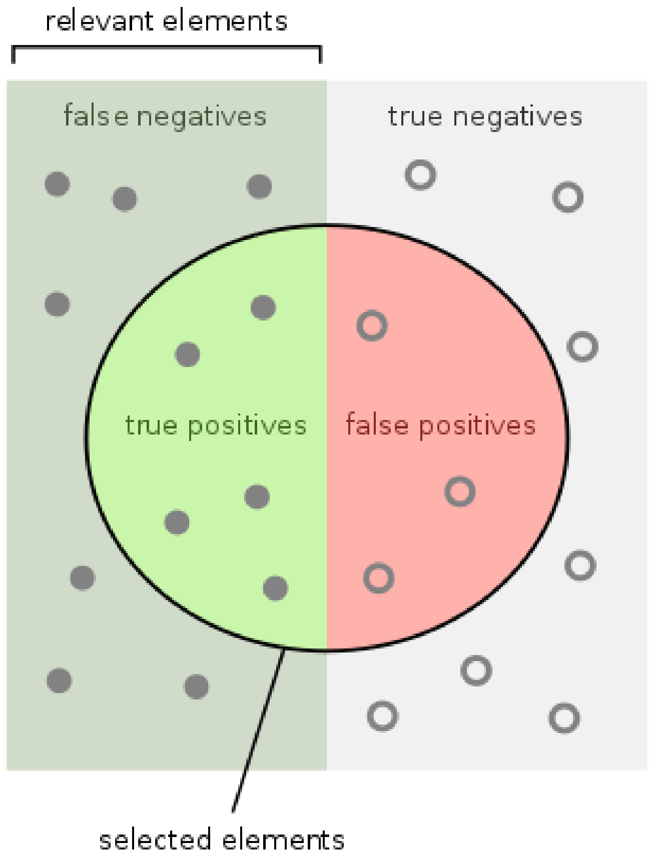


Признаки: цвет, форма

Решающее правило:
Груша желтая, овальная
Яблоко красное, круглое

Точность и ошибки бинарной классификации

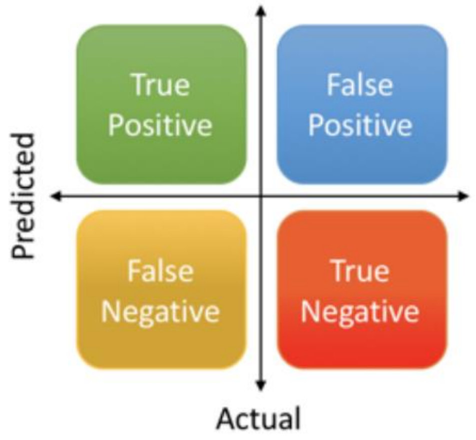
Ошибка первого рода – ложная тревога, false positive
Ошибка второго рода – пропуск события, false negative



$$\textbf{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Positive}}$$

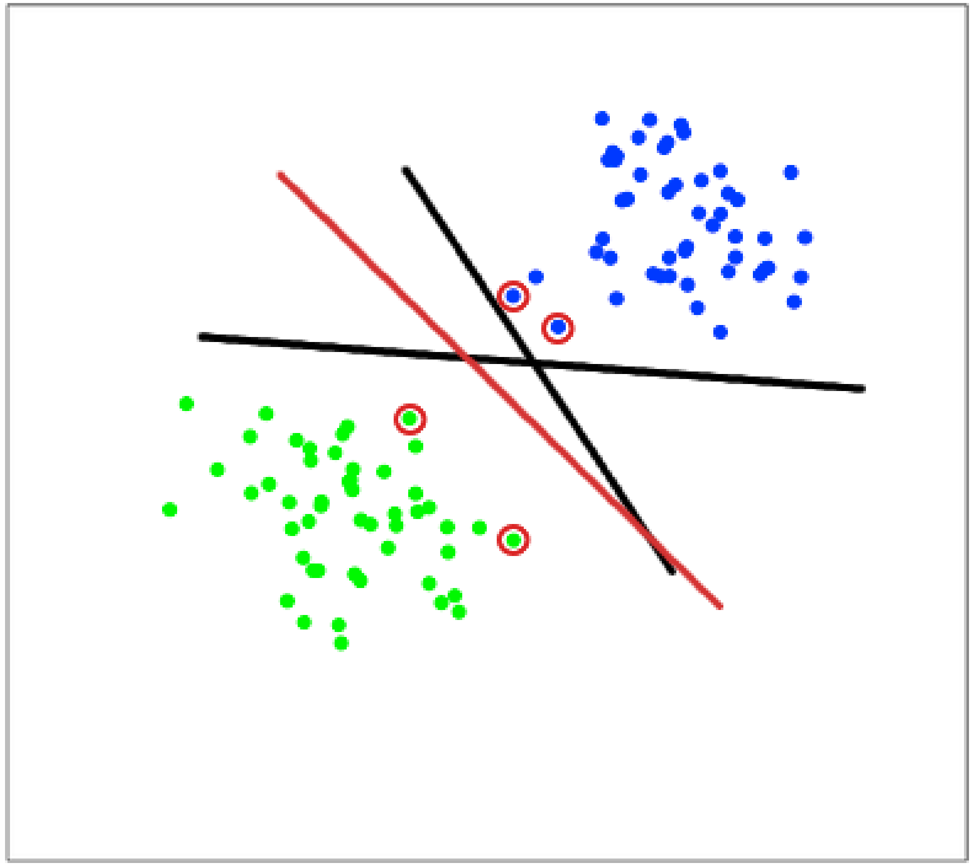
$$\textbf{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive + False Negative}}$$

$$\textbf{Accuracy} = \frac{\text{True Positive + True Negative}}{\text{Total}}$$



Простейшая задача классификации – метод опорных векторов

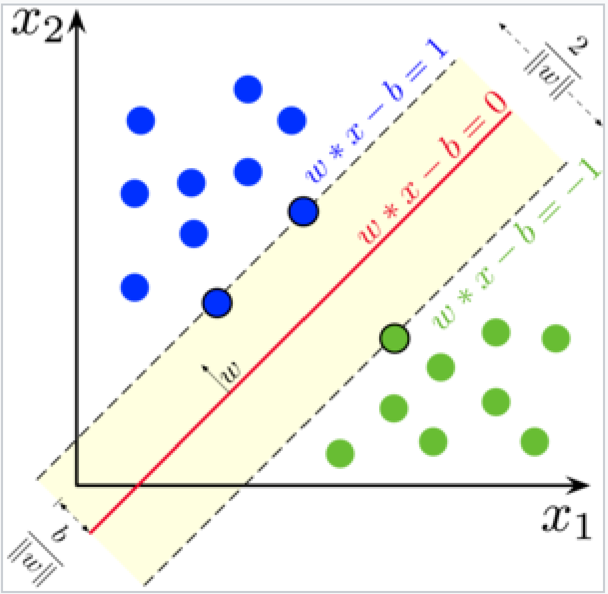
Линейное решающее правило бинарной классификации



Support Vector Machine

Support Vector Machine, Vapnik

Линейная разделимость



SVM основан на скалярном произведении:

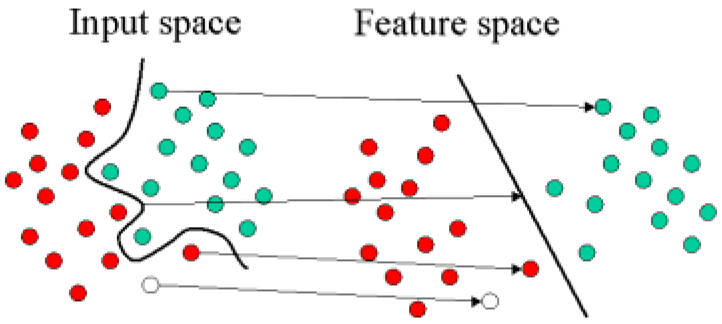
$$\underset{\mathbf{w}}{\text{Minimize}} \quad \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{subject to} \quad y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1$$

$$\mathbf{Y} \cdot (\mathbf{X}\mathbf{w} + b) \geq \mathbf{1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \cdot \left(\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1d} \\ x_{21} & x_{22} & \dots & x_{2d} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nd} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} + b \right) \geq \mathbf{1}^n$$

$$\boxed{w = \sum_{i=1}^{\ell} \lambda_i y_i x_i;}$$

Проблема: отсутствие разделимости



Решение проблемы разделимости:

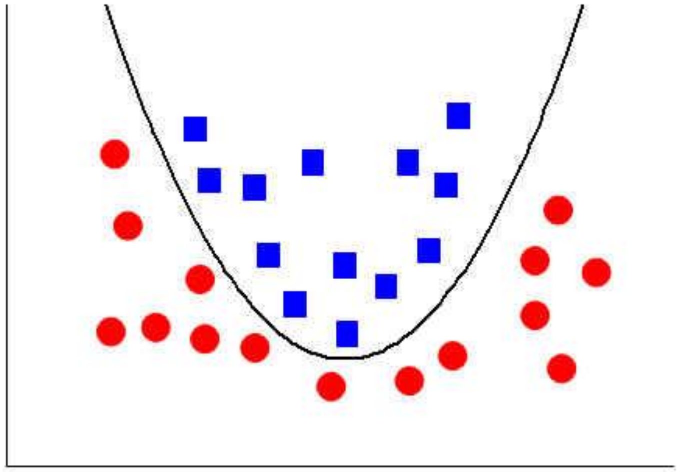
Ядерное сглаживание

Kernel regression

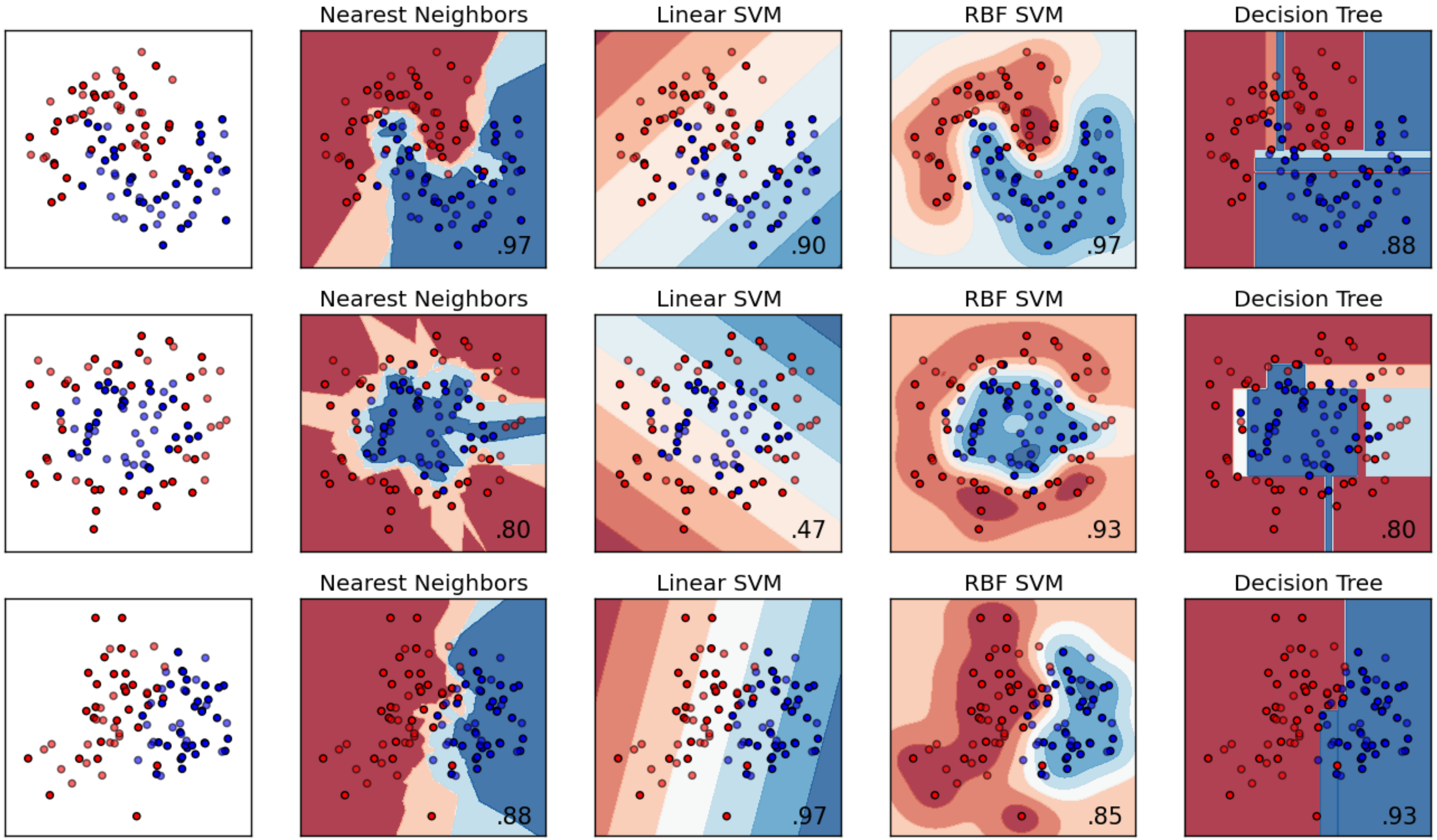
SVM основан на скалярном произведении **(x,y)**

С учетом ядра –

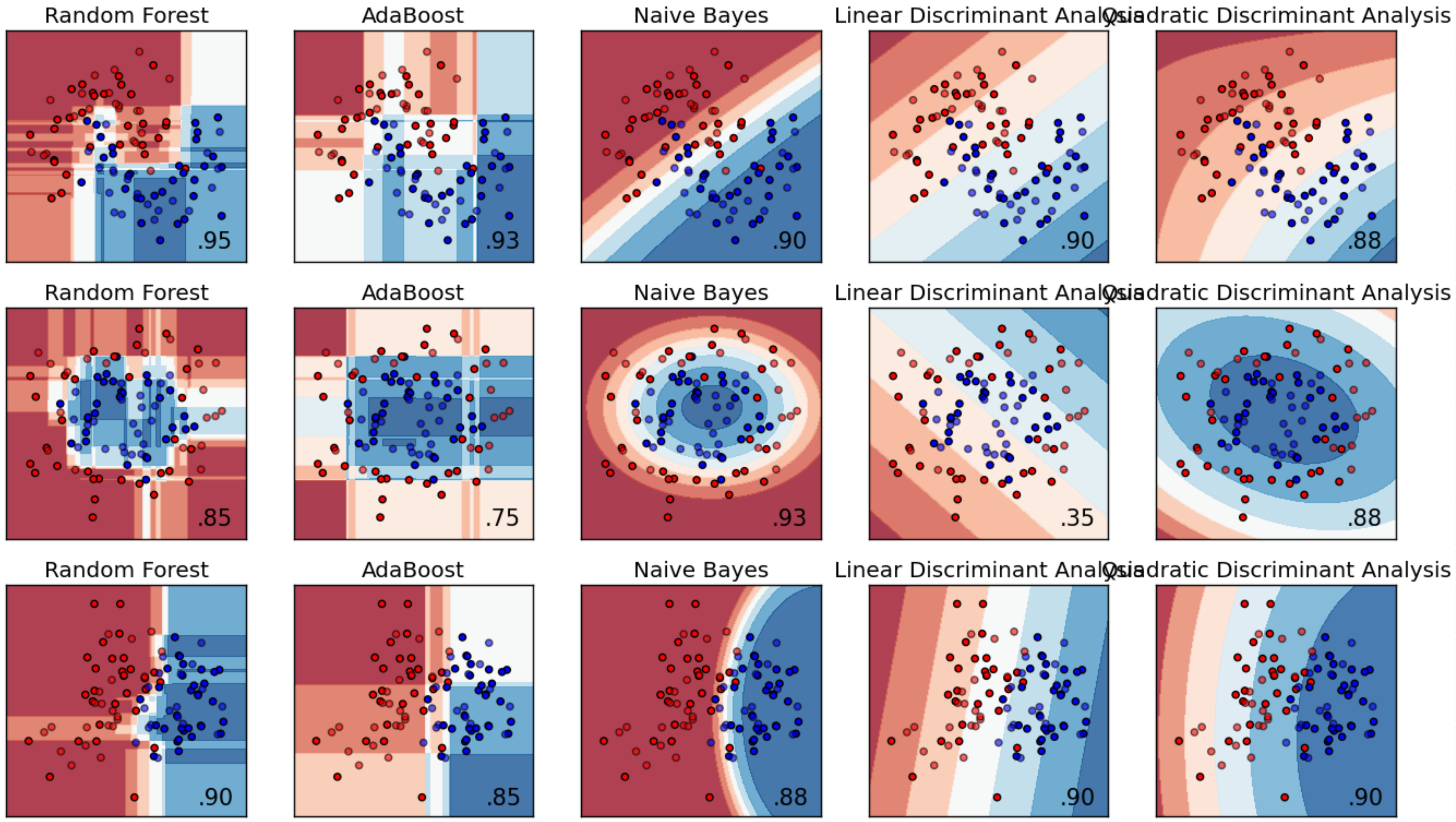
(X,Y) ~ (xKy)



Сравнение классификаторов

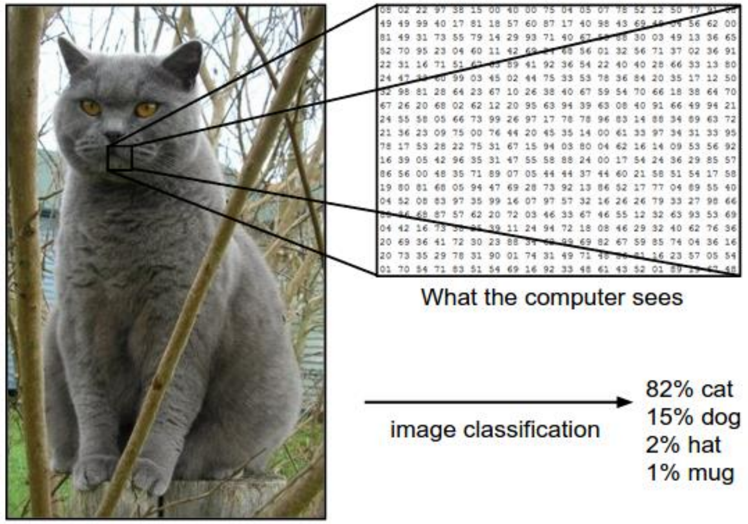


Сравнение классификаторов

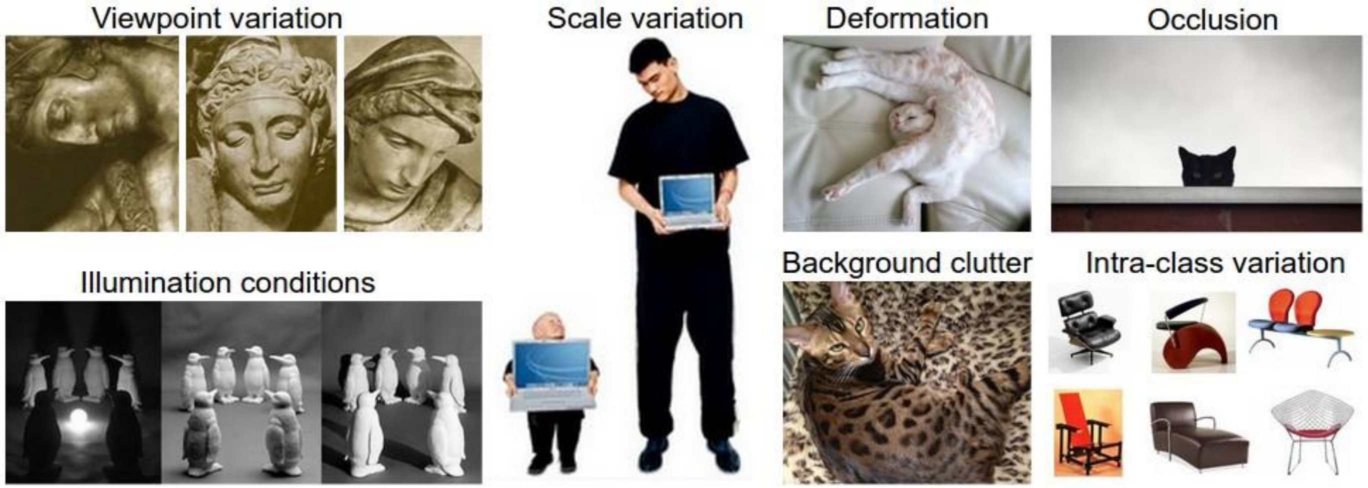


Отступление – анализ изображений

How the machine see the image?

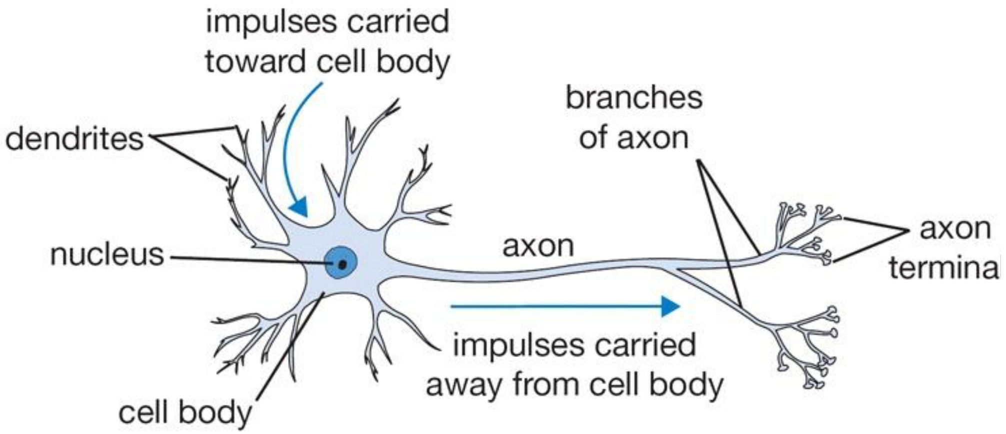


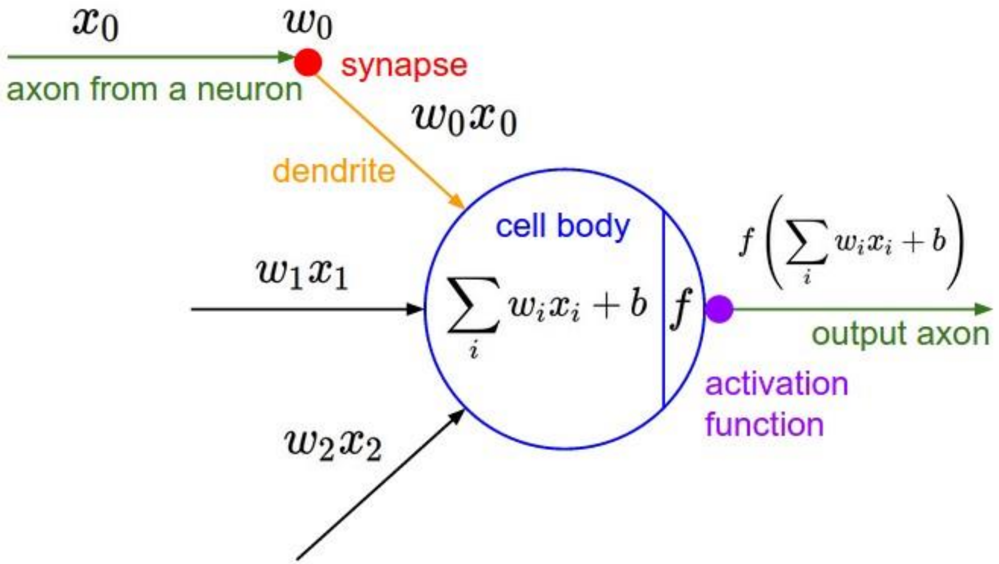
Основные проблемы при классификации изображений

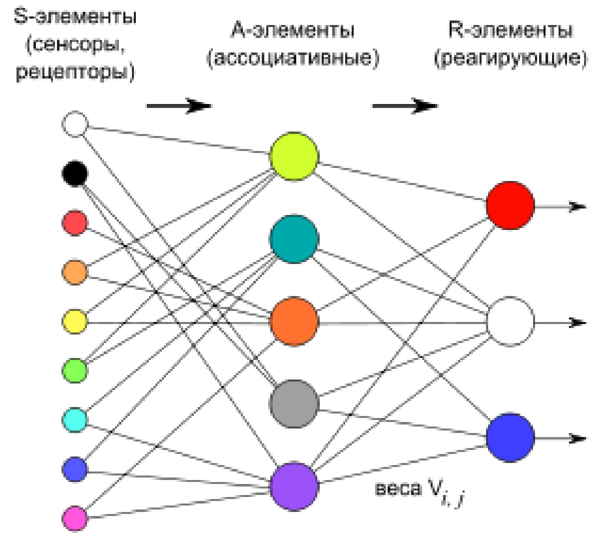


Нейронные сети

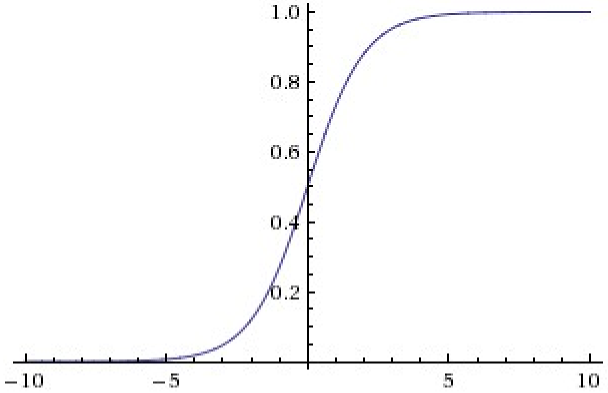
Нейробиологическая аналогия (неверная!)







Перцептрон, однослойный



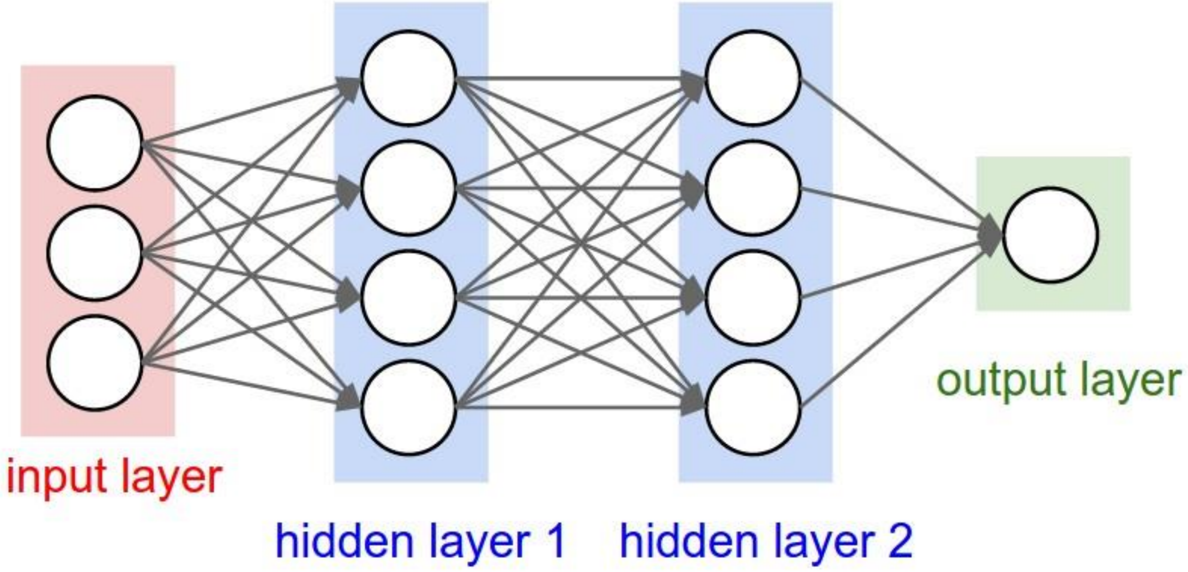
Сигмоидальная функция активации

Насколько мы близки к модели мозга?

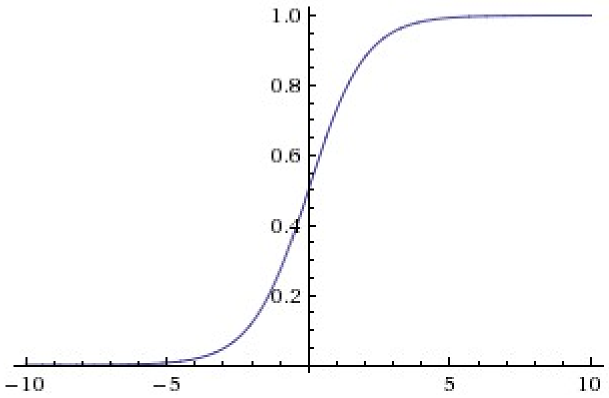
Для модели всего мозга проекту Blue Brain потребовалось бы 8.4 ГВт, проекту SpiNNaker – 0,2 ГВт, тогда как мощность Волжской ГЭС – 2,67 ГВт.

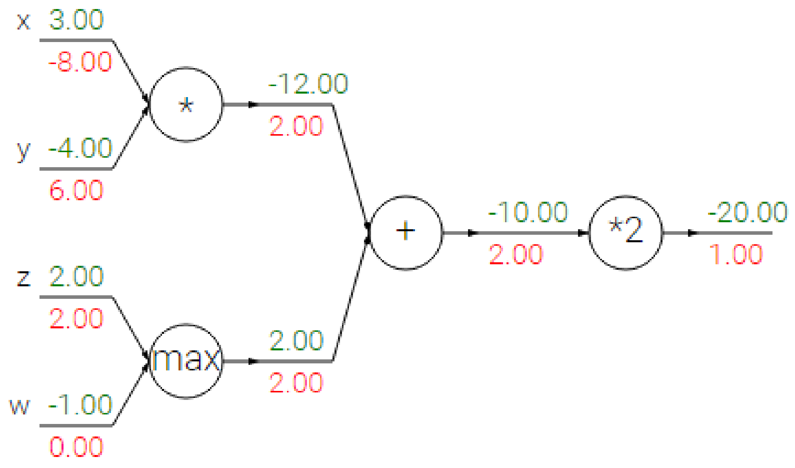
Multi-layer perceptron, Backpropagation algorithm

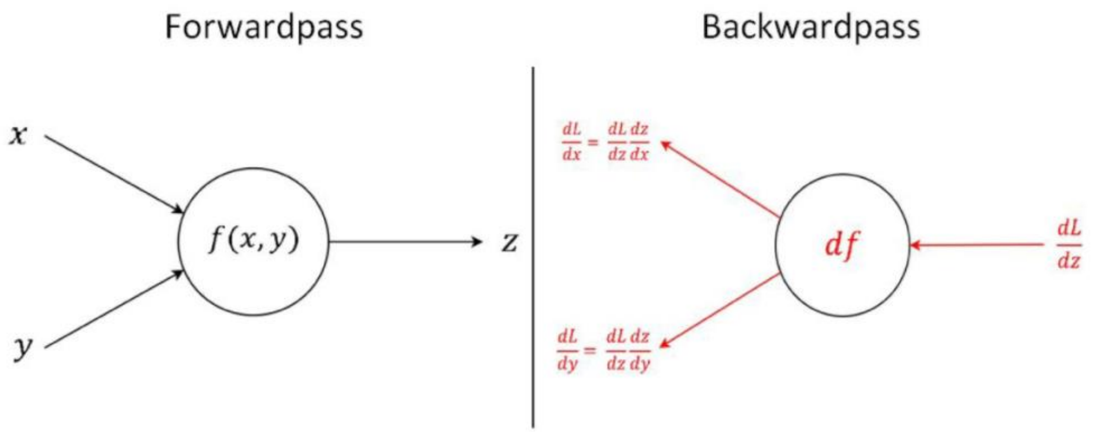
MLP



Функция активации







Stochastic Gradient Descent – Стохатический градиентный спуск

Minimizing of the cost function J(θ) over the data

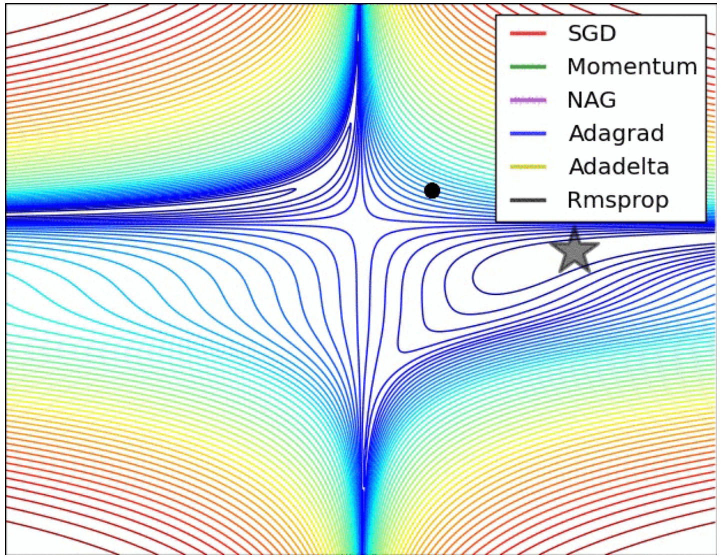
$\theta = \theta - \eta \cdot \nabla_\theta J(\theta).$ «Ванильный» градиентный спуск

$\theta = \theta - \eta \cdot \nabla_\theta J(\theta; x^{(i)}; y^{(i)}).$ Стохастический ГС η(λ) – learning rate

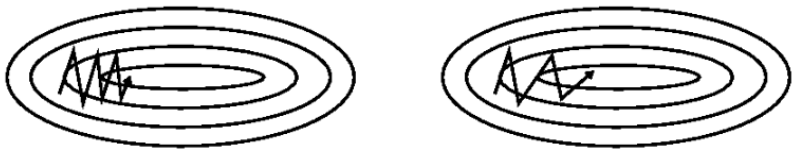
$\theta = \theta - \eta \cdot \nabla_\theta J(\theta; x^{(i:i+n)}; y^{(i:i+n)}).$ Mini-batch SGD – пакетный СГС

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta)$$
$$\theta = \theta - v_t$$

Модификации SGD учитывают анизотропию фазового пространства – Adam etc.



Momentum γ:



Регуляризация наше все!
- Weight decay
- Dropout
- Pruning – контрастирование
- Batch-norm

2. Weight penalty terms

L2 weight decay

$$E = \frac{1}{2}\sum_j (t_j - y_j)^2 + \frac{\lambda}{2}\sum_{i,j} w_{ji}^2$$
$$\Delta w_{ji} = \varepsilon \delta_j x_i - \varepsilon \lambda w_{ji}$$

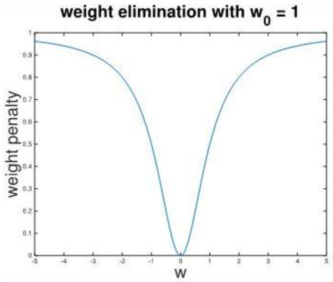
L1 weight decay

$$E = \frac{1}{2}\sum_j (t_j - y_j)^2 + \frac{\lambda}{2}\sum_{i,j} |w_{ji}|$$
$$\Delta w_{ji} = \varepsilon \delta_j x_i - \varepsilon \lambda \, \mathrm{sign}(w_{ji})$$

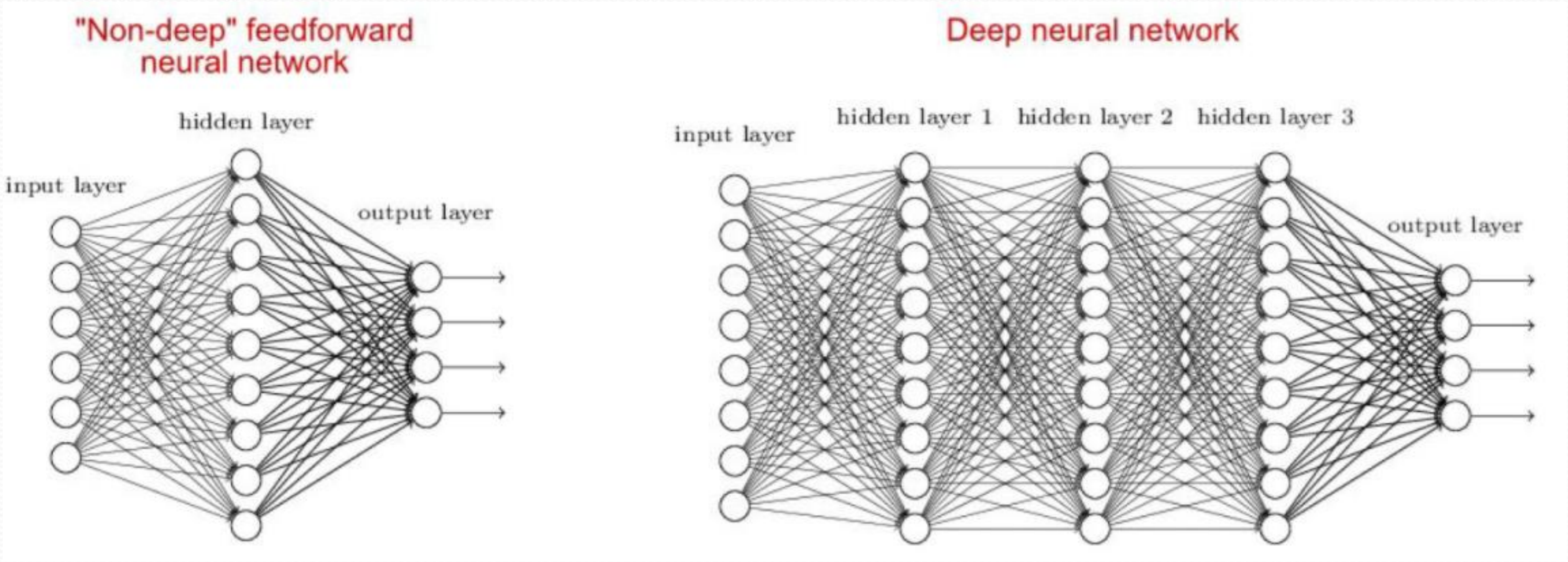
weight elimination

$$E = \frac{1}{2}\sum_j (t_j - y_j)^2 + \frac{\lambda}{2}\sum_{i,j} \frac{w_{ji}^2 / w_0^2}{1 + w_{ji}^2 / w_0^2}$$

See Reed (1993) for survey of 'pruning'



Shallow vs Deep Network

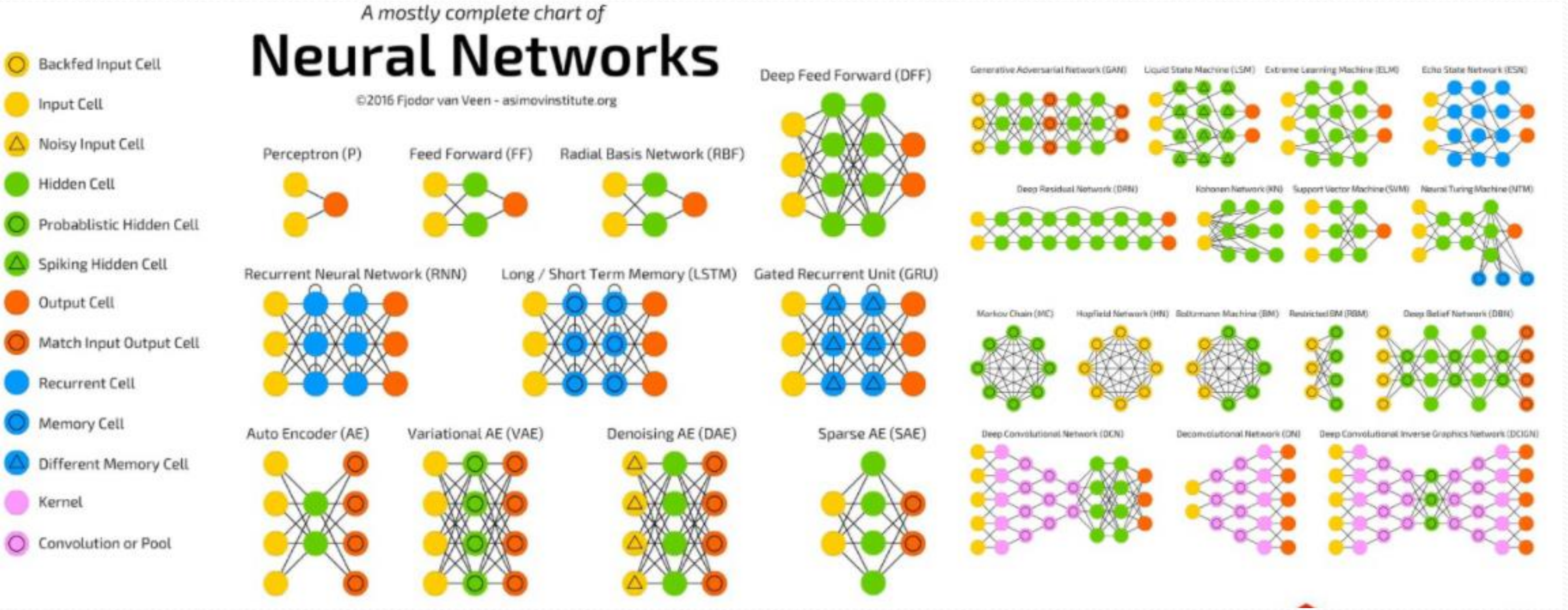


Почему обучение *глубокое*, а не *глубинное*?

Пожалуй лучший вводный курс от Стэнфорда: http://cs231n.github.io/

Пожалуй лучшая книжка на русском: С. И. Николенко, А. Кадурин, Е. В. Архангельская, Глубокое обучение. Погружение в мир нейронных сетей

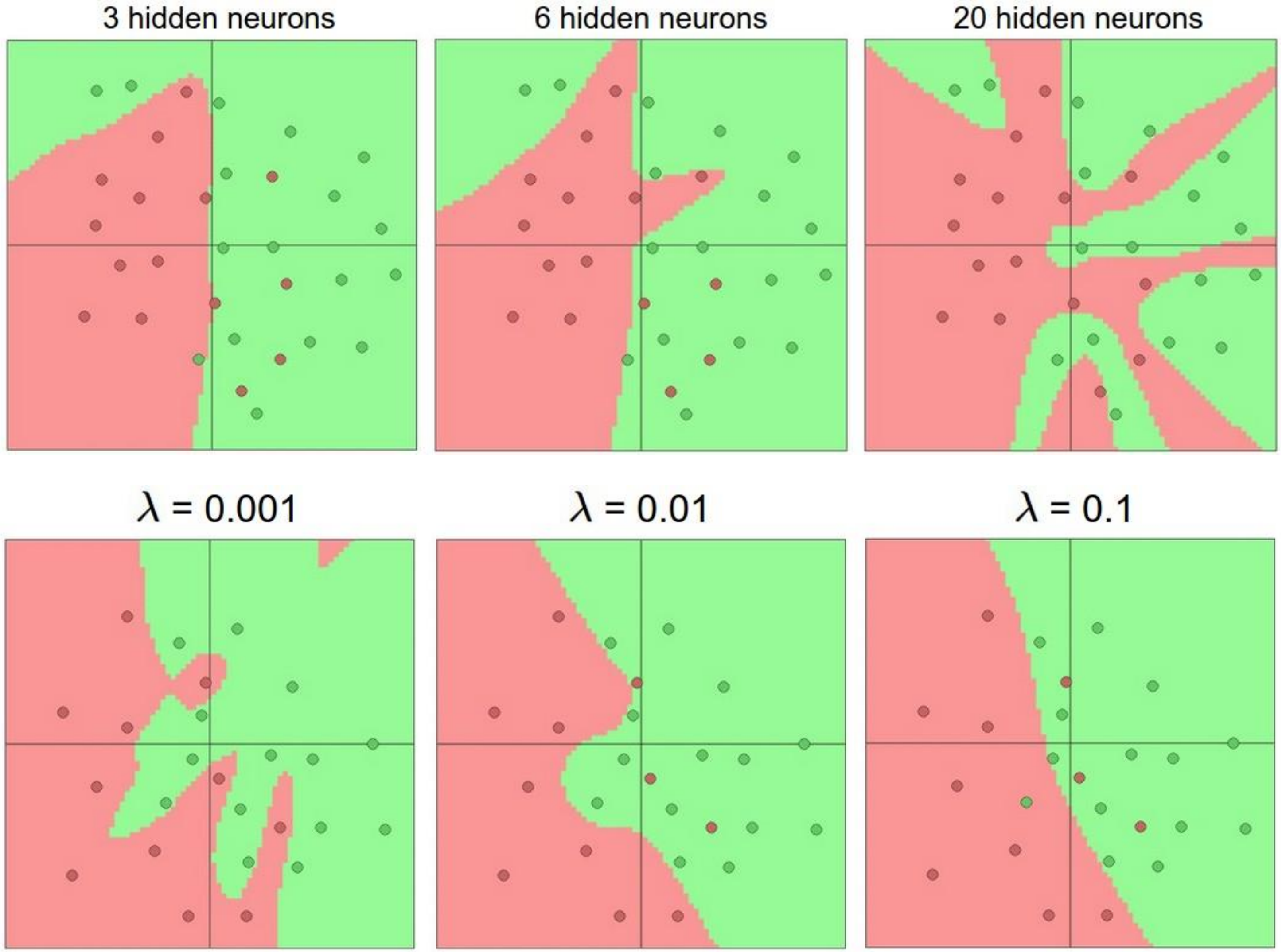
Complete Chart of Neural Networks



https://towardsdatascience.com/the-mostly-complete-chart-of-neural-networks-explained-3fb6f2367464

Проблемы классических нейронных сетей

Недообучение и переобучение



Проблемы

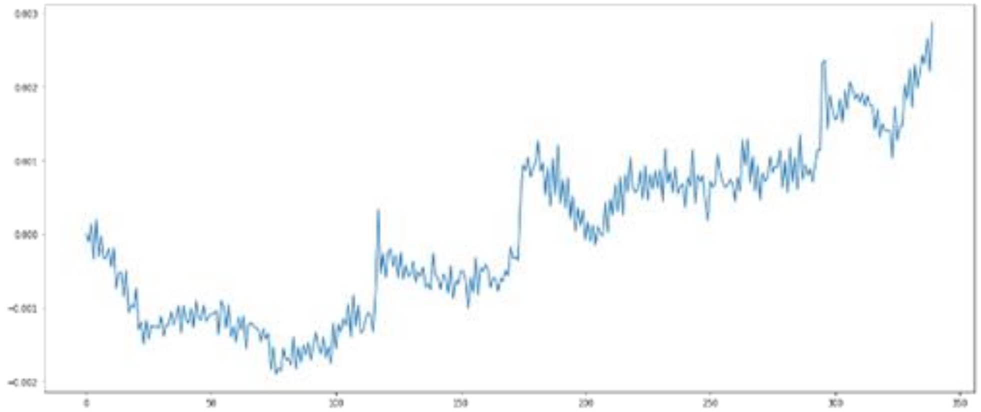
1. Выбор структуры
2. Инженерия признаков
3. Overfit
4. Dead gradients
5. Интерпретация

Решения

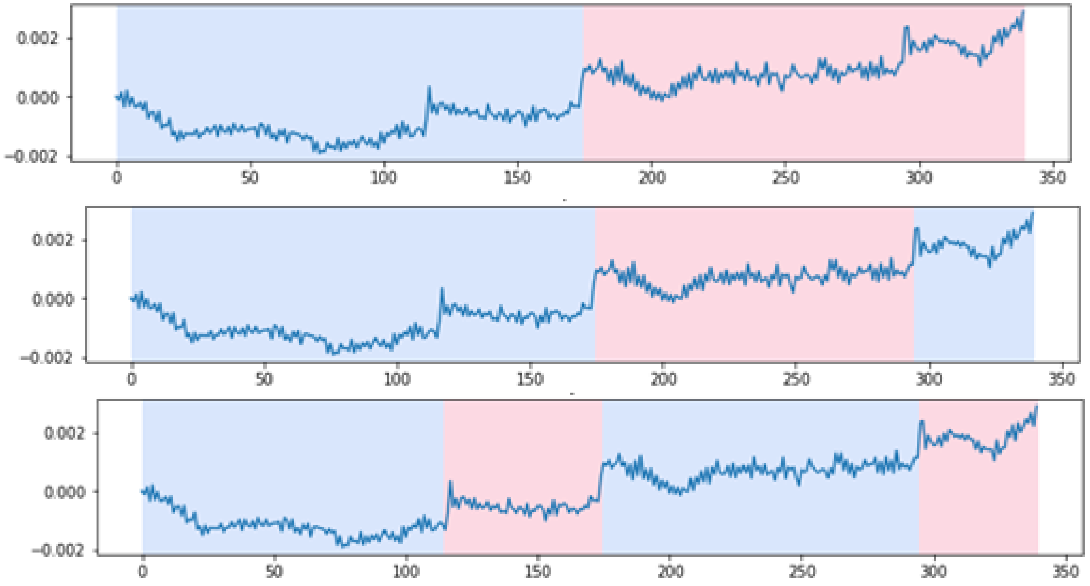
1. Learning rate
2. Регуляризация
3. Контрастирование

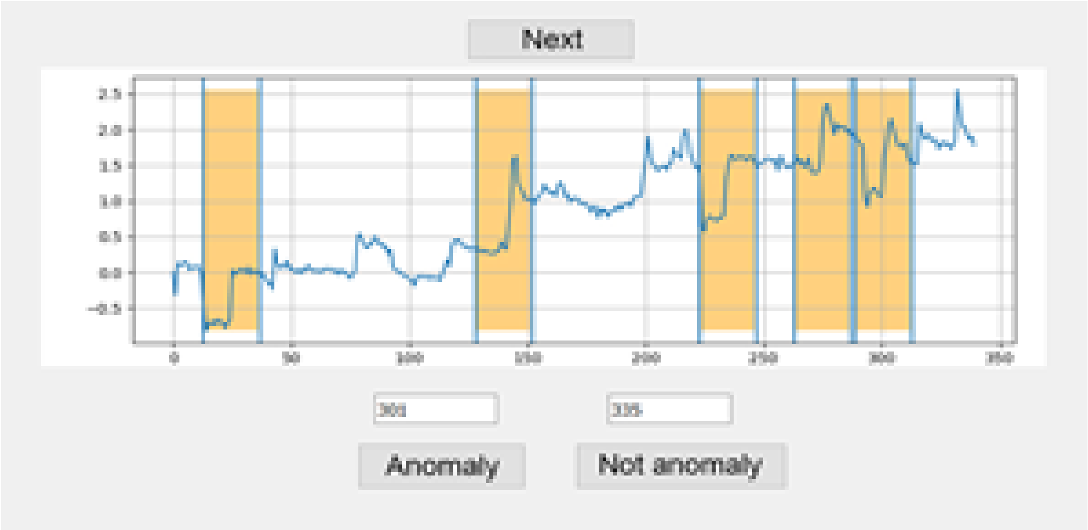
Пример задачи. Классификация аномалий

Задача. Детектирование аномалий типа «скачок», step



Временной ряд со скачками





Размечаем вручную – получаем датасет, **обучаем классификатор**

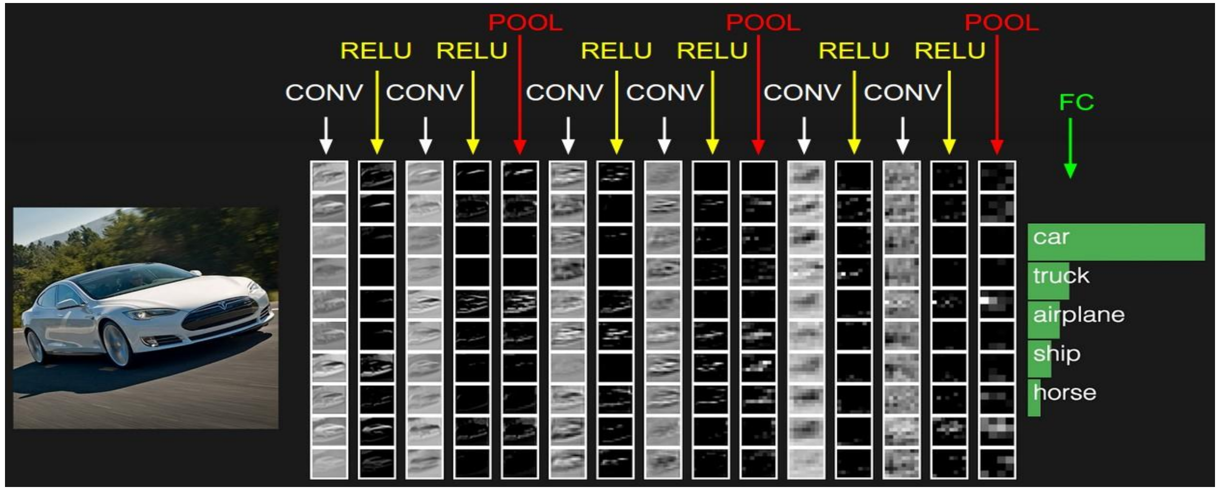
Статистическое детектирование – плохо (https://github.com/deepcharles/ruptures)

Convolutional networks CNN, Сверточные сети

Convolutional Neural Nets, CNN

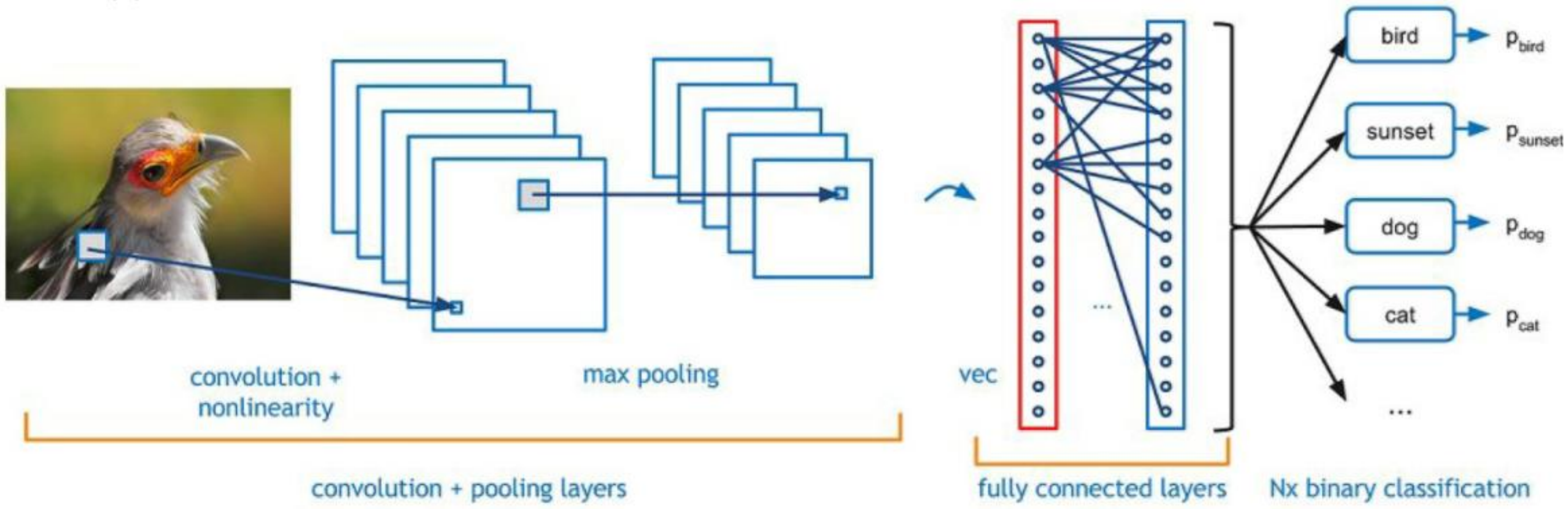
LeNET 5, 1988, Y. LeCun

AlexNet, 2012, A. Krizhevsky, I. Sutskever and G. Hinton



Yann LeCun

Geoffrey Hinton



Сверточные сети и GPU

1989 G Cybenko
Теорема об универсальной аппроксимации

Approximation by superpositions of a sigmoidal function - Springer Link
https://link.springer.com/article/10.1007/BF02551274 - Перевести эту страницу
автор: G Cybenko - 1989 - Цитируется: 10688 - Похожие статьи

1998 Yann LeCun
сверточные сети

ieeexplore.ieee.org › document - Перевести эту страницу
Gradient-based learning applied to document recognition ...
Gradient-based learning applied to document recognition. ... A new **learning** paradigm, called graph transformer networks (GTN), allows such multimodule systems to be trained globally using **gradient-based** methods so as to minimize an overall performance measure. Two systems for online handwriting **recognition** are described.
автор: Y Lecun - 1998 - Цитируется: 28105 - Похожие статьи

2007 – Выход NVIDIA CUDA,

2009 – Google отказывается от нейронных сетей

2012 – AlexNet

[PDF] ImageNet Classification with Deep Convolutional Neural Networks
https://papers.nips.cc/.../4824-imagenet-classification-with-de... ▾ Перевести эту страницу
автор: A Krizhevsky - 2012 - Цитируется: 34232 - Похожие статьи

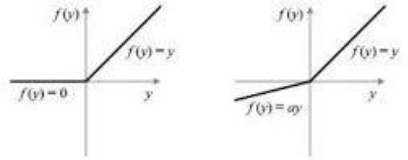


Figure 1. ReLU vs. PReLU. For PReLU, the coefficient of the negative part is not constant and is adaptively learned.

Delving Deep into Rectifiers: Surpassing Human-Level Performance .
https://arxiv.org › cs ▾ Перевести эту страницу
автор: K He - 2015 - Цитируется: 3856 - Похожие статьи

IEEE CVPR Cite Score: 3.23 (2012), 6.19 (2015), 18.18 (2018)

CNN layers, слои СНС

Выделение признаков + классификация

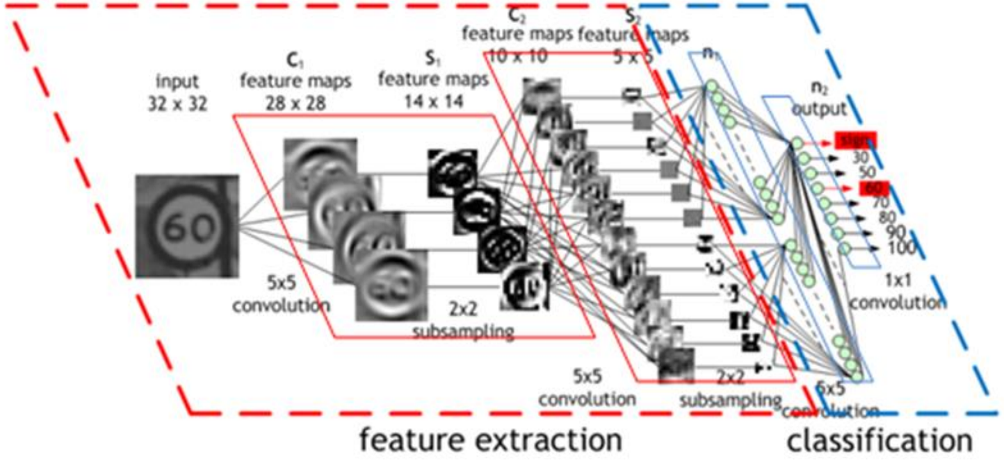
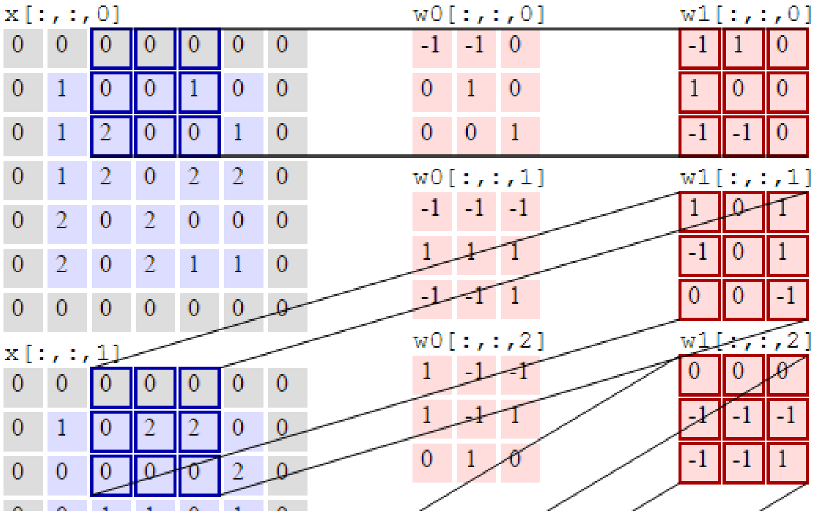


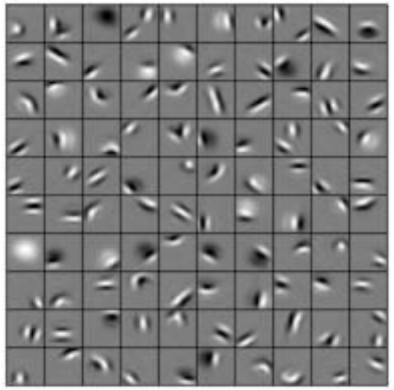
Иллюстрация работы сверточного слоя



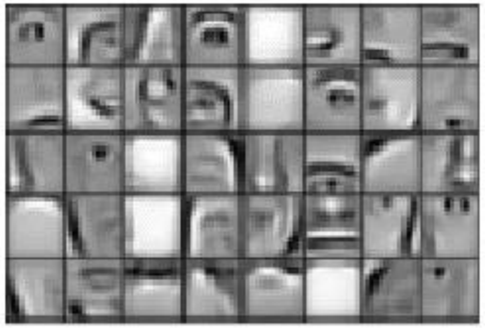
Решаемые проблемы

- Переобучение
- Привыкание к данным
- Выделение признаков перестало быть искусством

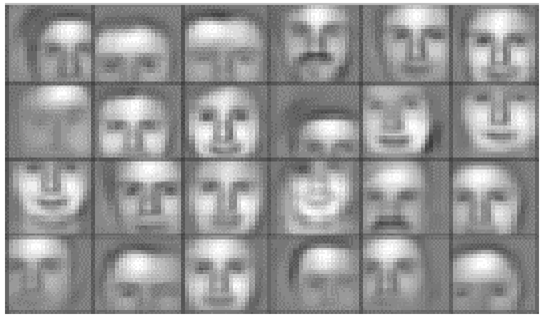
Feature maps, карты признаков:



У входного слоя



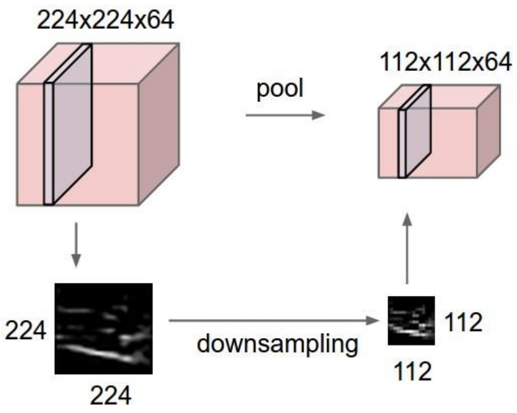
В середине

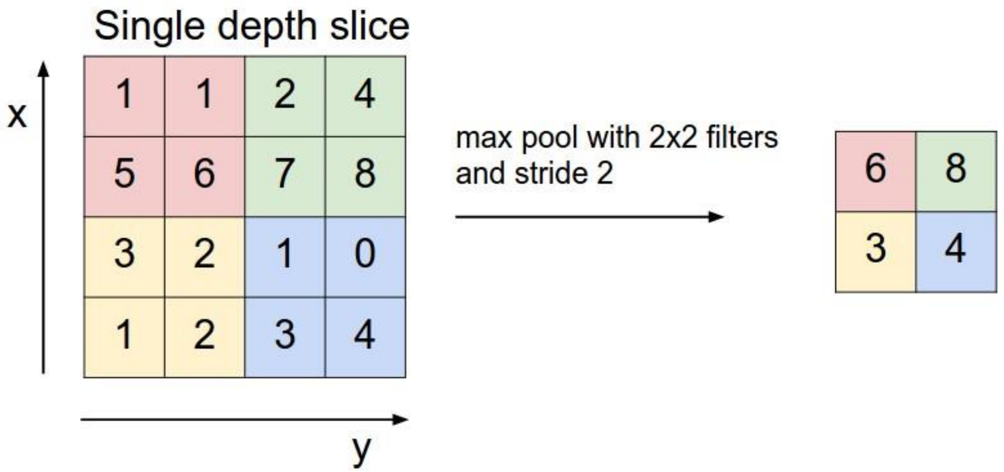


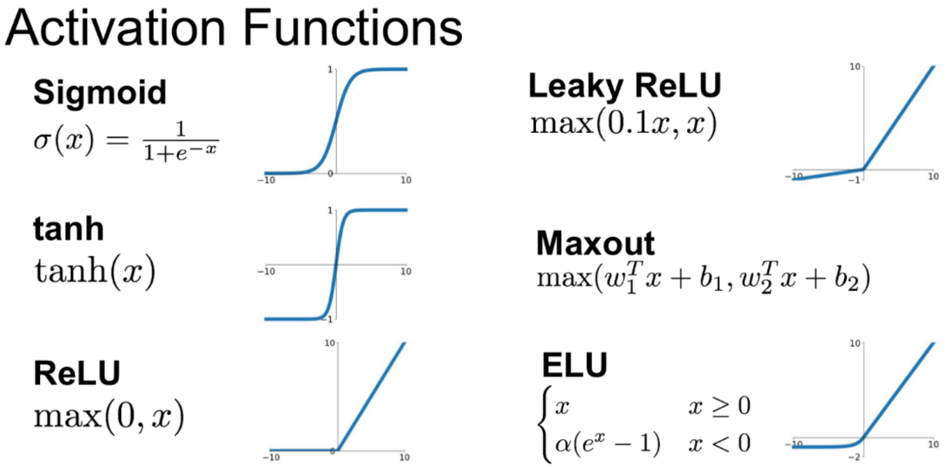
У выходного слоя

CNN layers, слои СНС

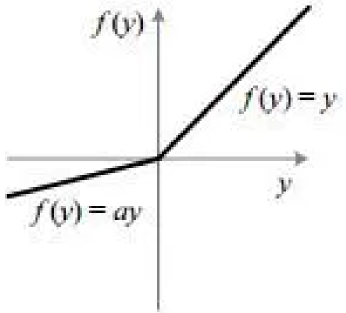
Pooling



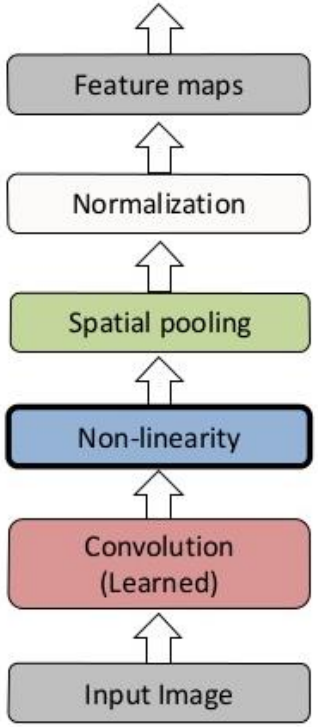




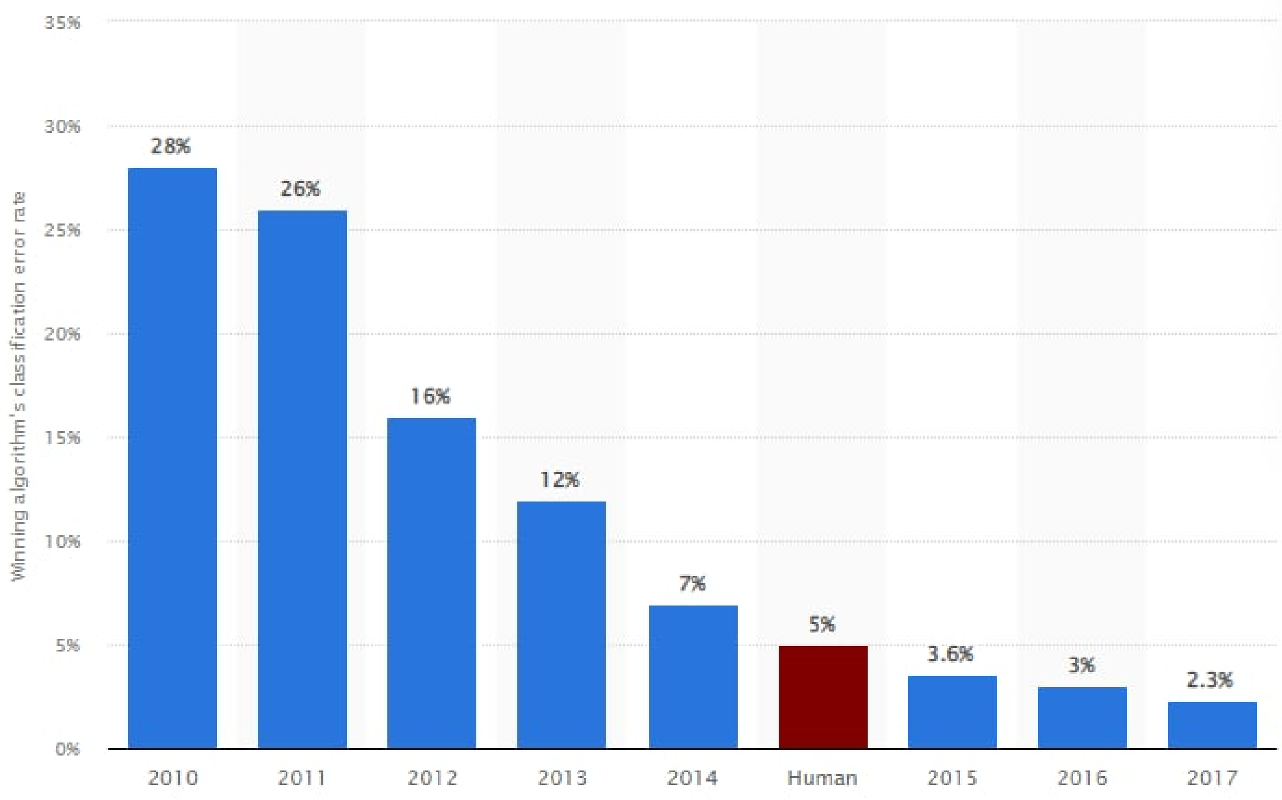
PReLU:

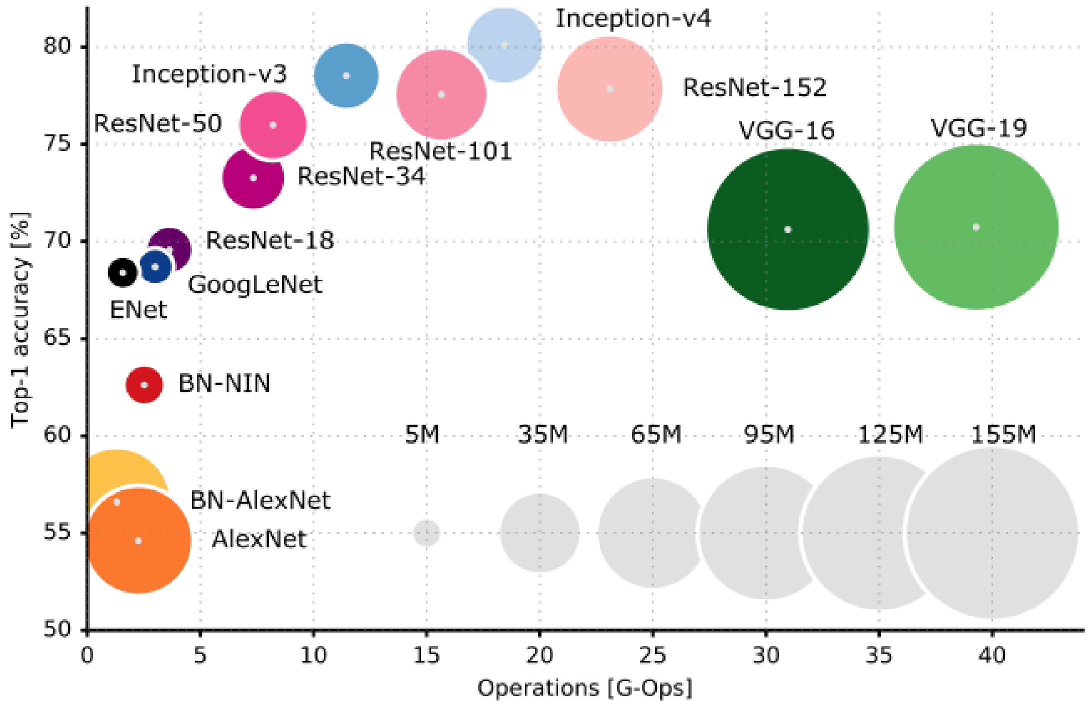


Обработка картинки сетью



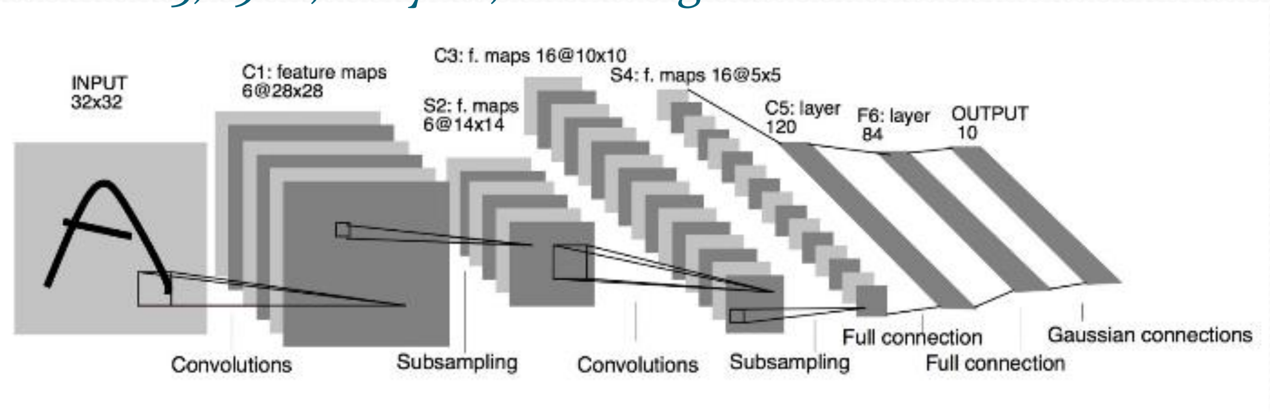
Deep and Accurate



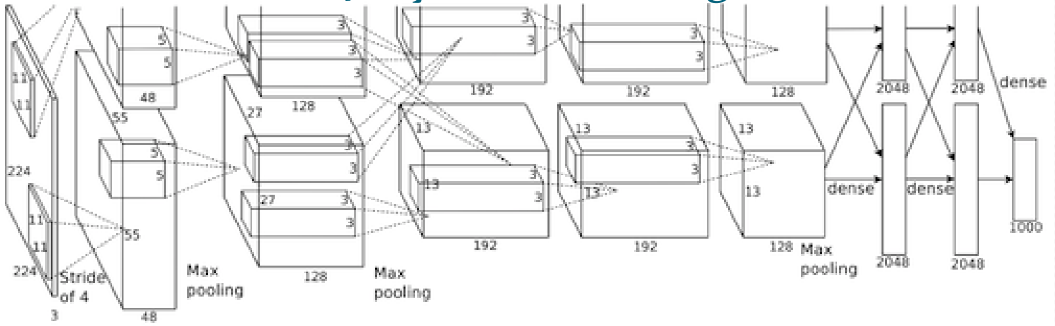


Typical Architectures 1

LeNet5, 1988, 8 layers, 60K weights



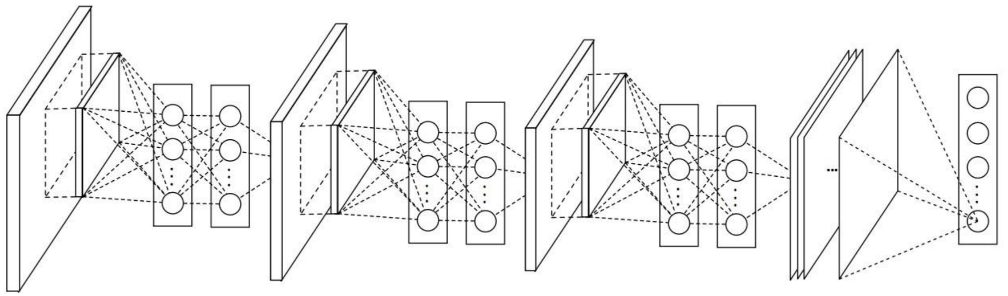
AlexNet, 2012, 7 layers, 60 M weights



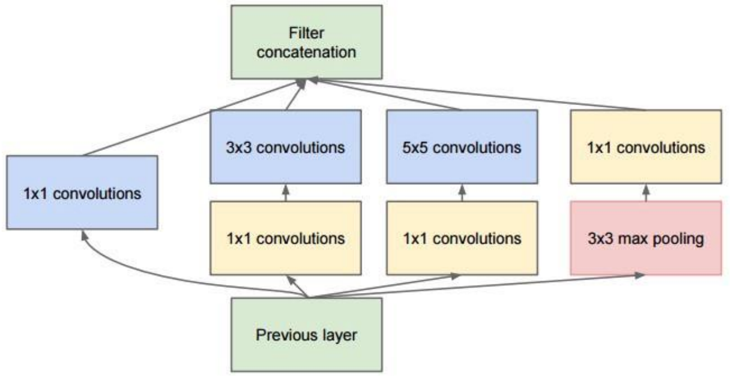
VGG, 2014, 16 layers, 138 M

Typical Architectures 2

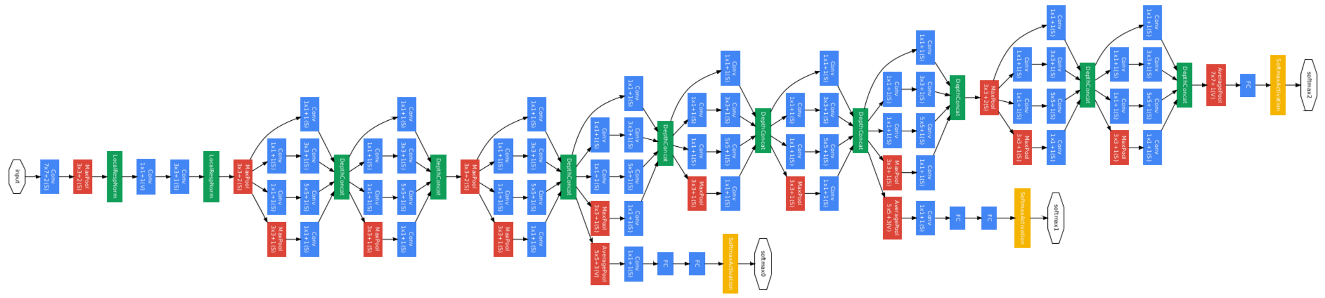
Network in network, 2013



Inception, 2014

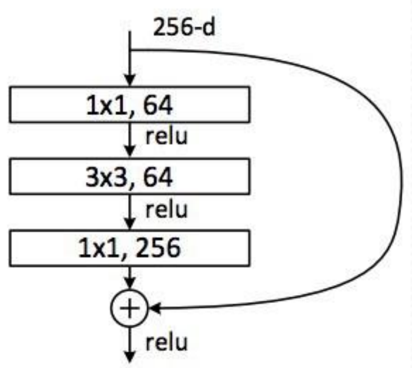


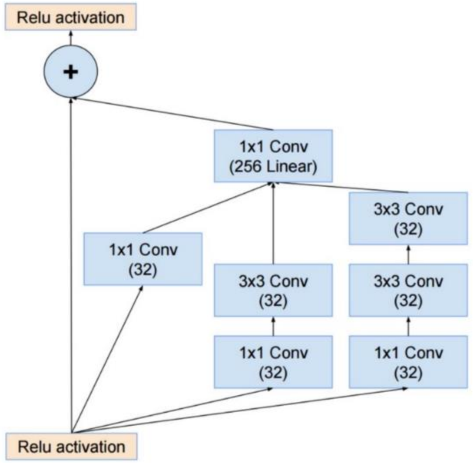
GoogleNet, 2014, 19 layers, 4 M weights



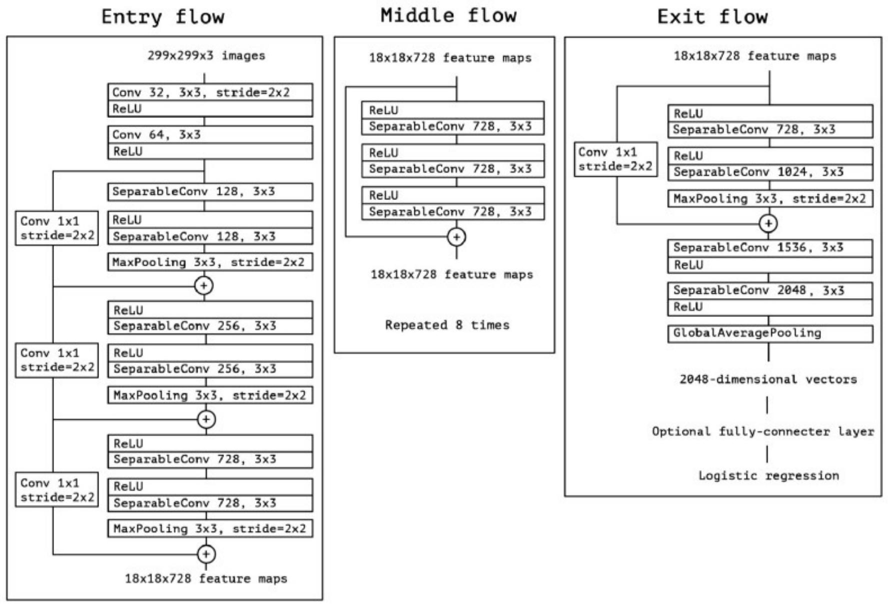
Typical Architectures 3

ResNet, Inception with Resnet module



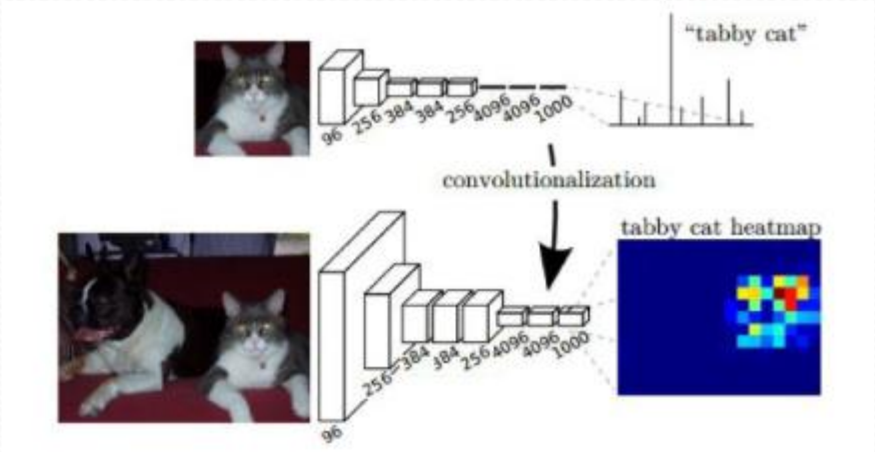


Xception

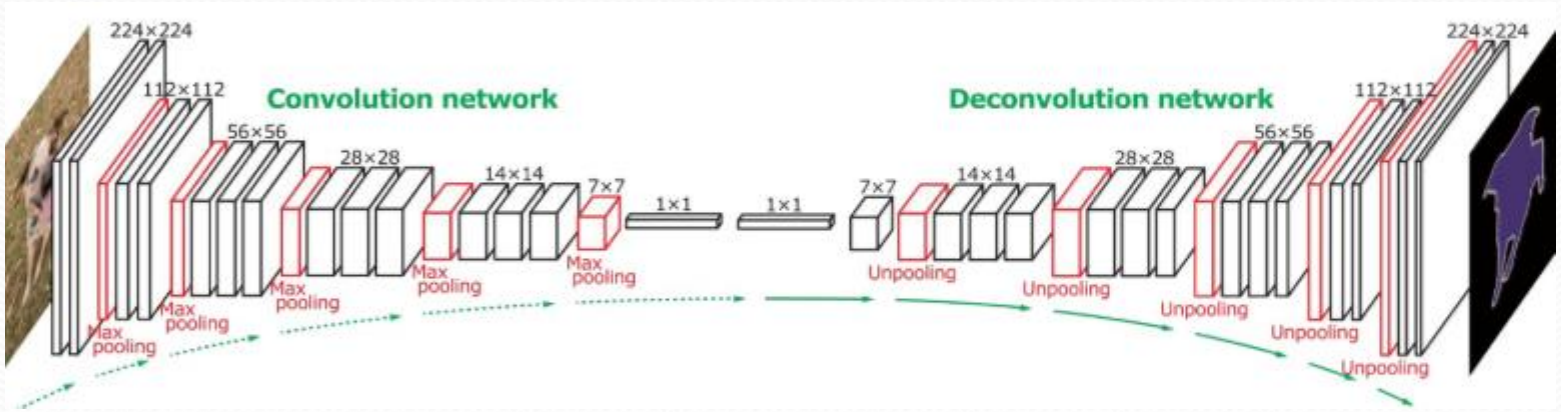


Typical Architectures 4

Fully convolutional network, FCN

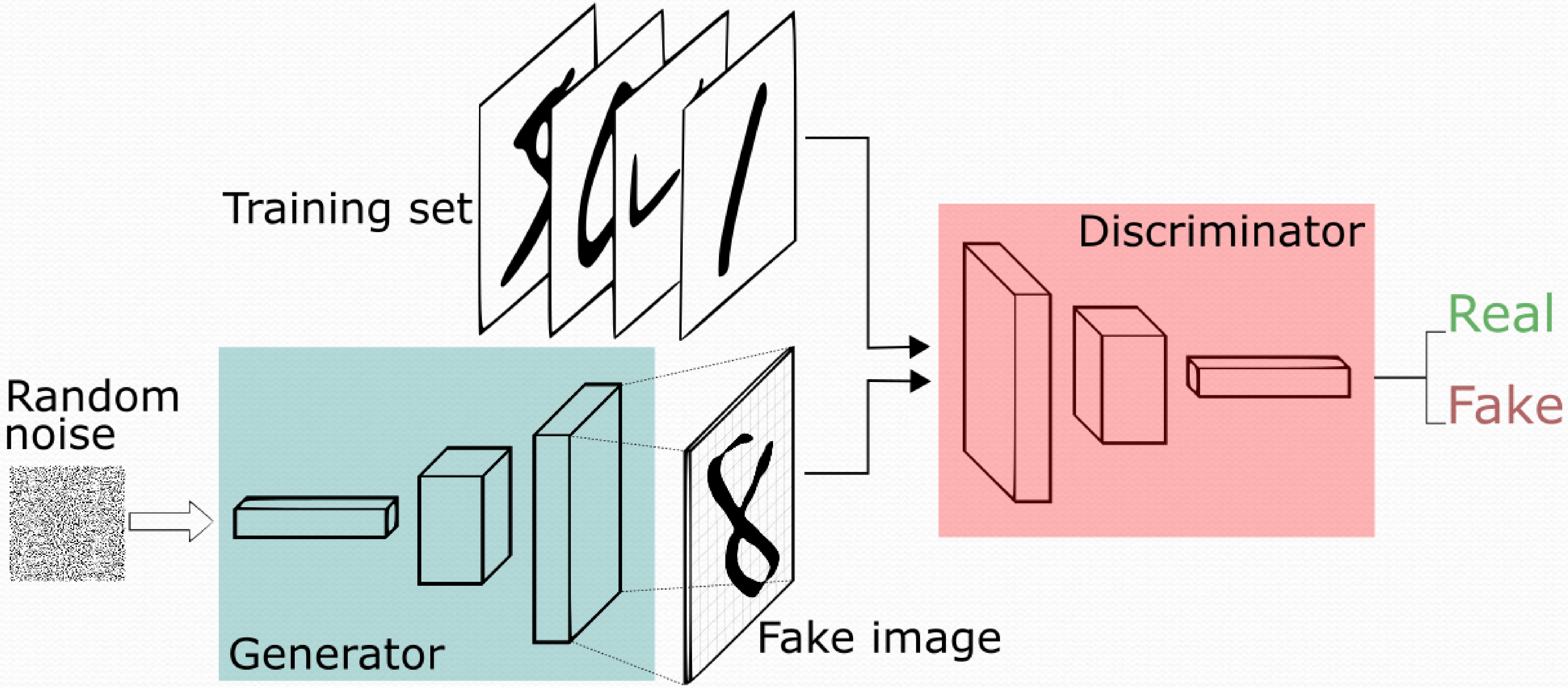


Deconvolutional network, Deconv



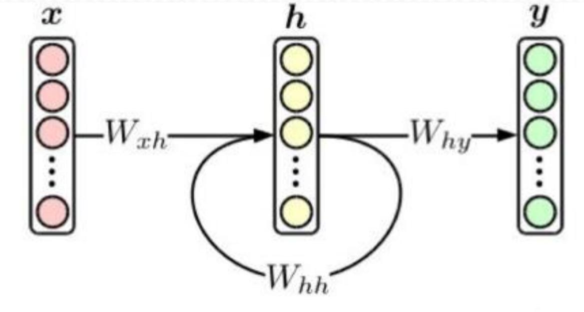
Typical Architectures 5

Генеративно-состязательные сети
Generative adversarial network

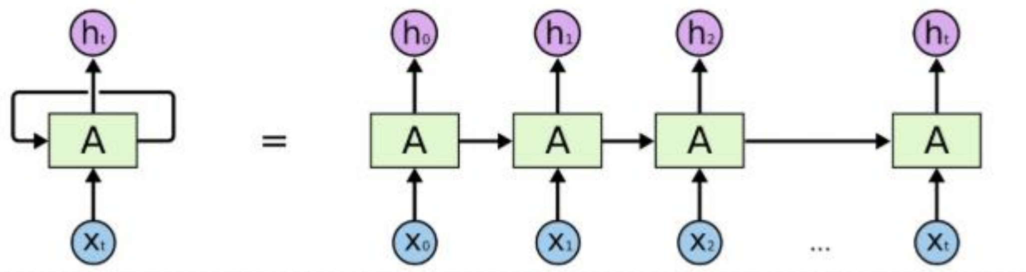


Typical Architectures - RNN

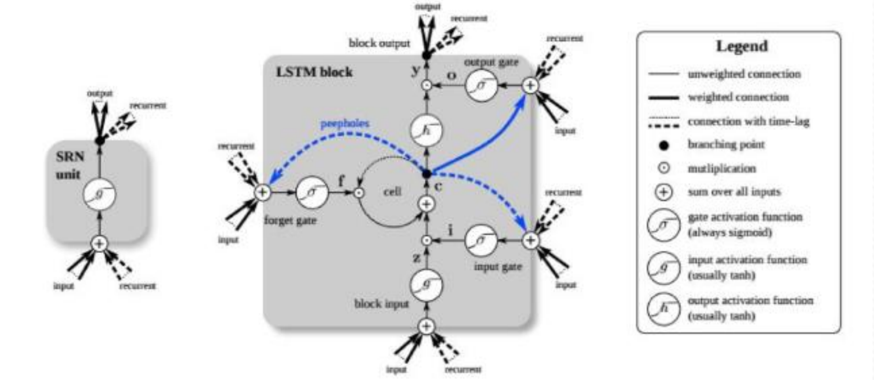
Recurrent NN, Turing complete!



Backpropagation through time

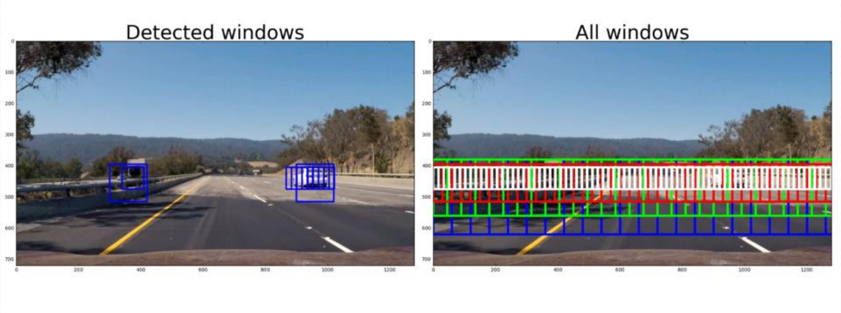


Long-Short Term Memory - LSTM

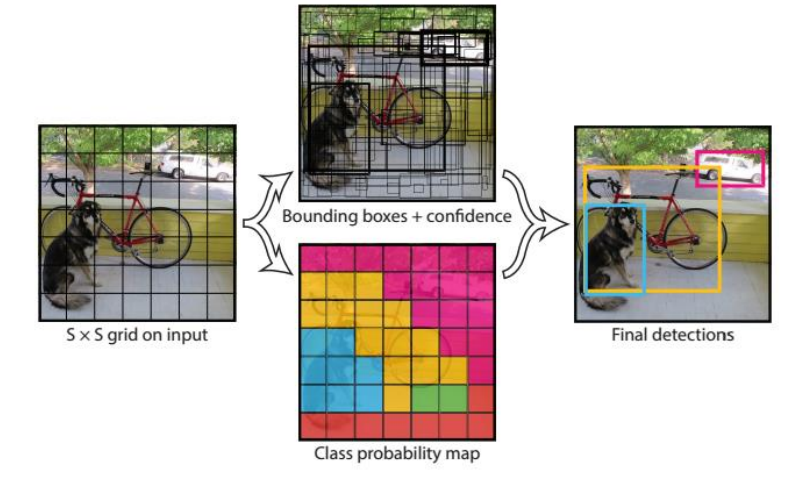


Untypical Solutions for typical problems

Fast Detection

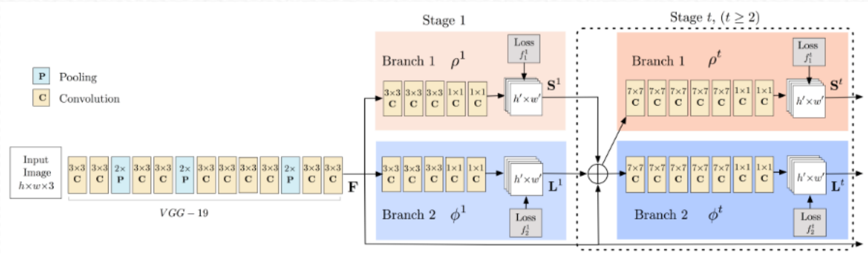


Sliding window detection -NO



YOLO - You Only Look Once – Yes!

Tracking



Real time pose tracking, CVPR 17

Openpose

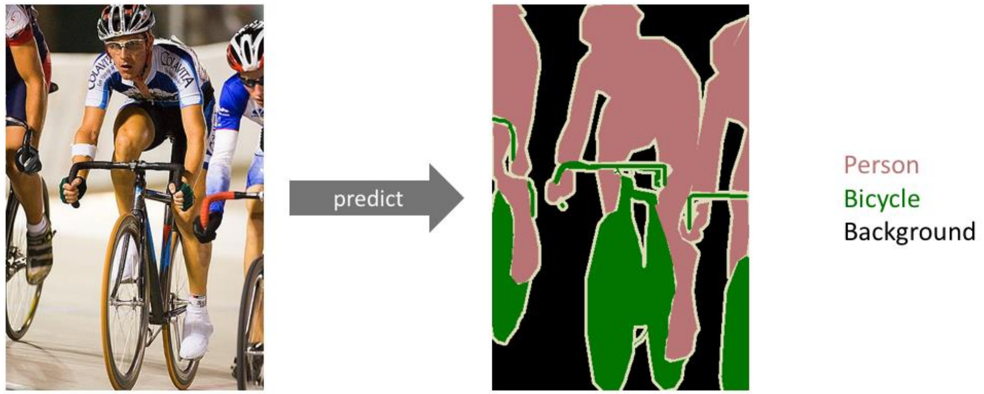
Вычислительная фотография –

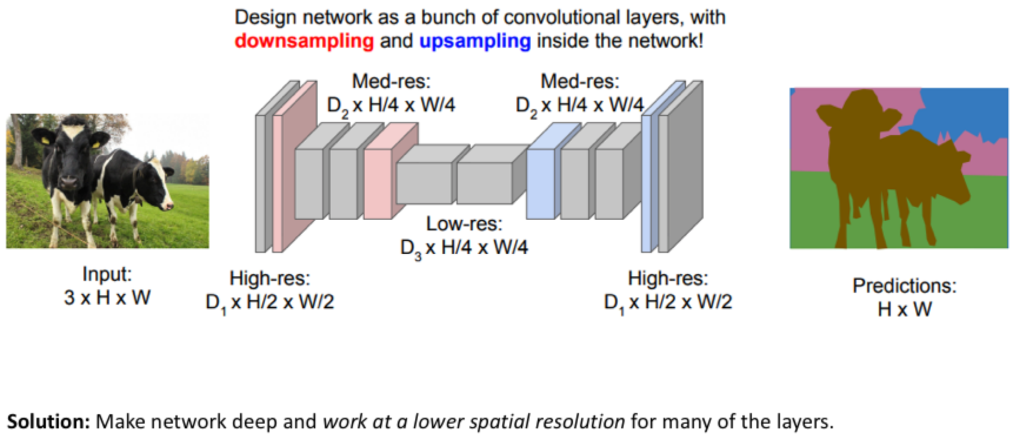
- Стекинг фото
- Сверхразрешение
- Съемка в темноте

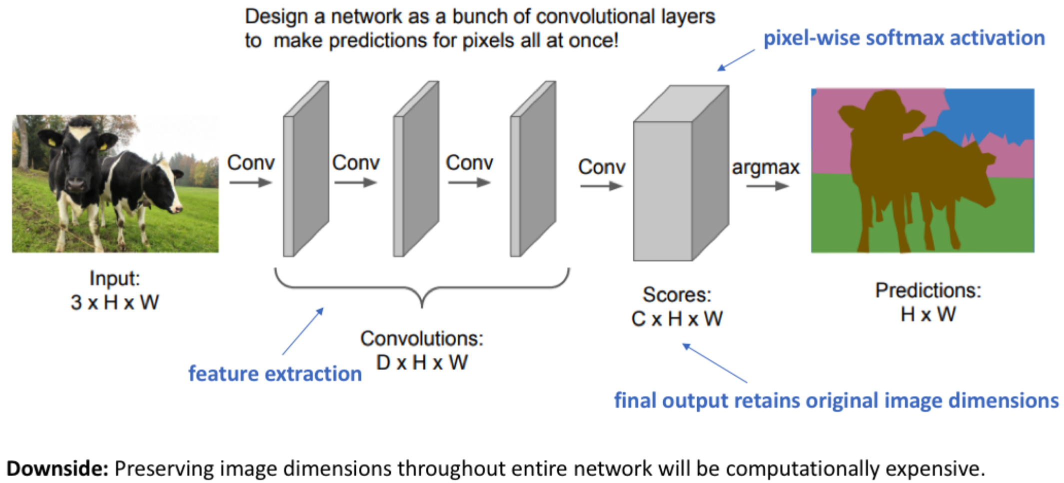
Light.co

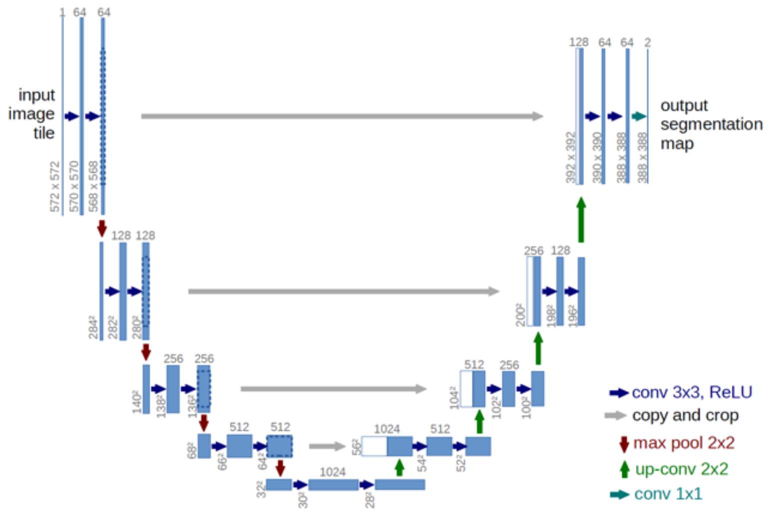
Unet сегментация и не только

Задача сегментации



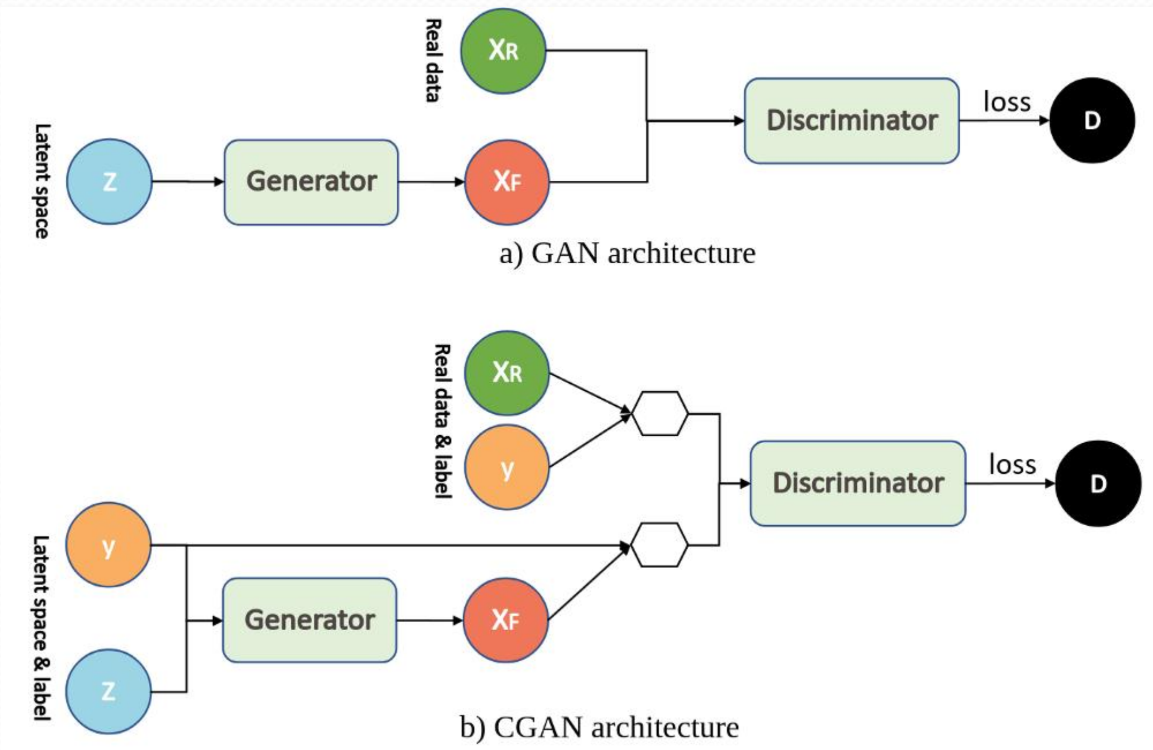


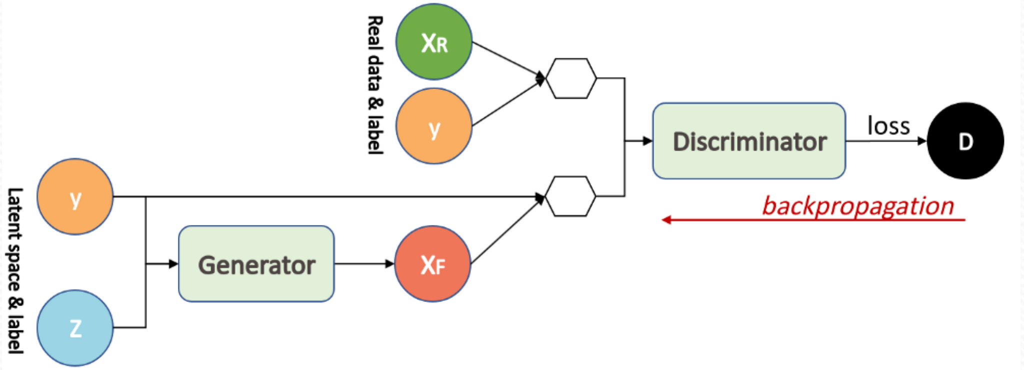


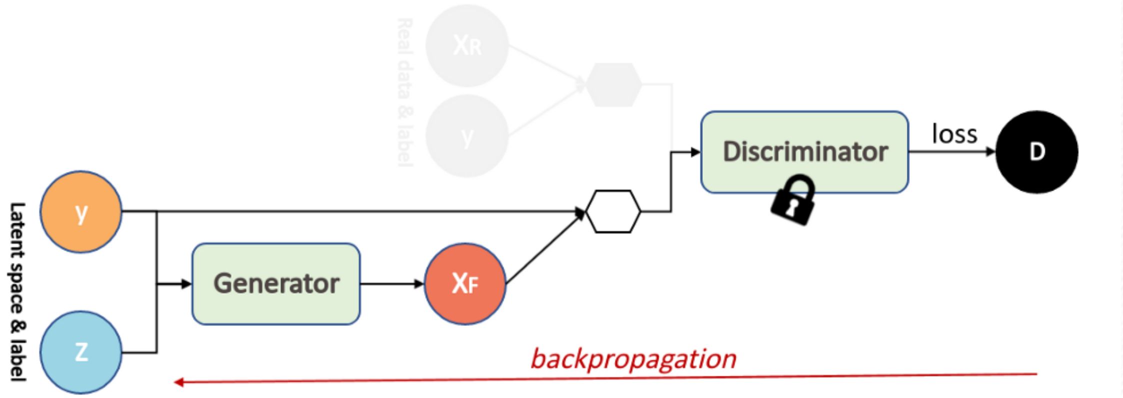


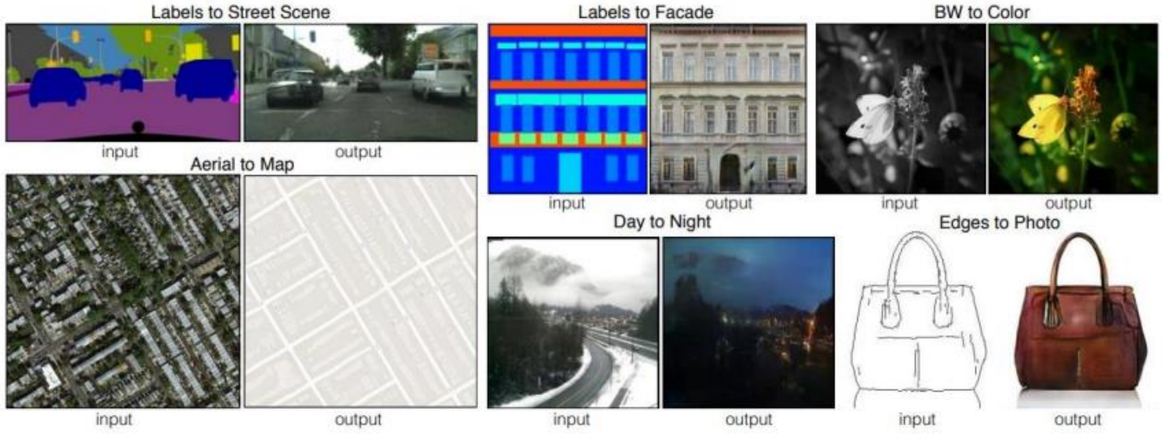
Добавим skip-connections – получим Unet!

Генеративно-состязательные сети (GAN)









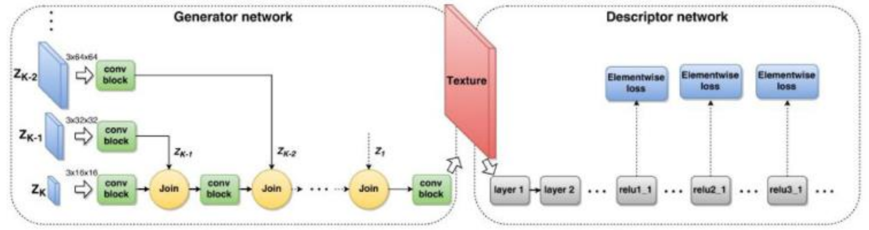
http://www1.idc.ac.il/toky/seminarIP-18/Presentations/10b_raaz.pdf

Image-to-Image Translation, Philip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros (Nov 2017)

Challenges

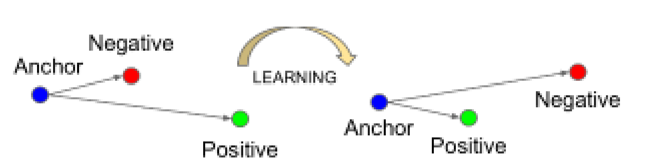
Solved

1. Self driving
2. Image enhancement
3. Single Image Super Resolution
4. Image annotation
5. Generator network



Content Style Pastiche

6. FaceNet



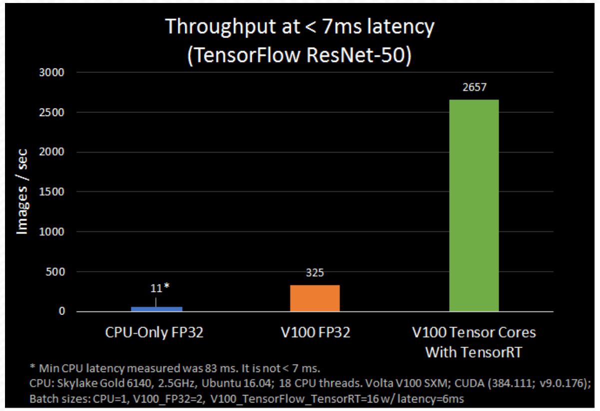
Unsolved / partially unsolved

1. Multi-object tracking
2. Fast target tracking
3. Medical image segmentation
4. Symmetry detection
5. Hyperspectral image processing
6. Fast inference for multiply videostreams
7. One shot learning

Hardware

Training:

- Nvidia GPU
- 1080 is 3 times better than 980
- No datacenter deployment feature
- Half precession
- Tensor cores
- Volta v100



GPU PERFORMANCE COMPARISON

	P100	V100	Ratio
DL Training	10 TFLOPS	120 TFLOPS	12x
DL Inferencing	21 TFLOPS	120 TFLOPS	6x
FP64/FP32	5/10 TFLOPS	7.5/15 TFLOPS	1.5x
HBM2 Bandwidth	720 GB/s	900 GB/s	1.2x
STREAM Triad Perf	557 GB/s	855 GB/s	1.5x
NVLink Bandwidth	160 GB/s	300 GB/s	1.9x
L2 Cache	4 MB	6 MB	1.5x
L1 Caches	1.3 MB	10 MB	7.7x

Inference (high performance inference)

Nvidia Jetson TX2 – 15 вт, 85 г, 1 TOп.

Intel Movidus Myriad X – 8x8 мм, 1 гр, 1 вт, 4 Топ

Huawei Kirin 970, 980

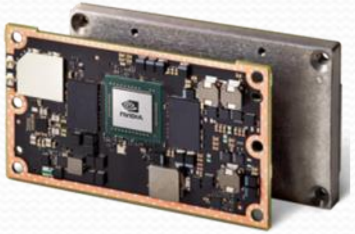
Jetson Nano!

Special

Google TPU

IBM TrueNorth

Module NeuroMatrix



Software Frameworks

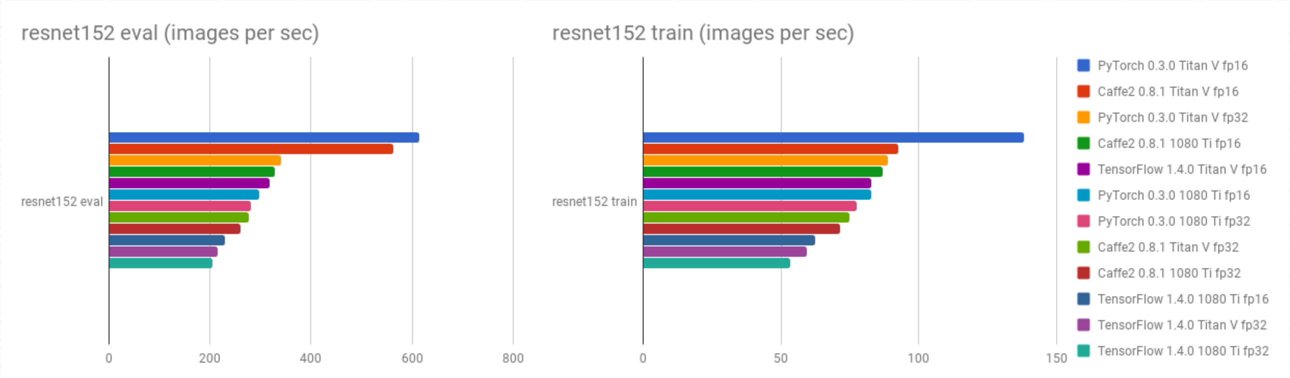
Training:
Tensorflow
Caffe
Torch
CNTK
MXNET

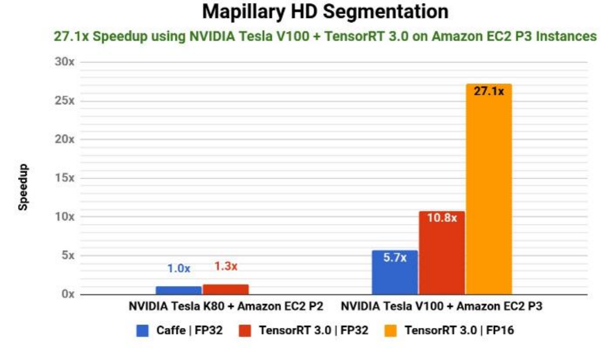
Keras

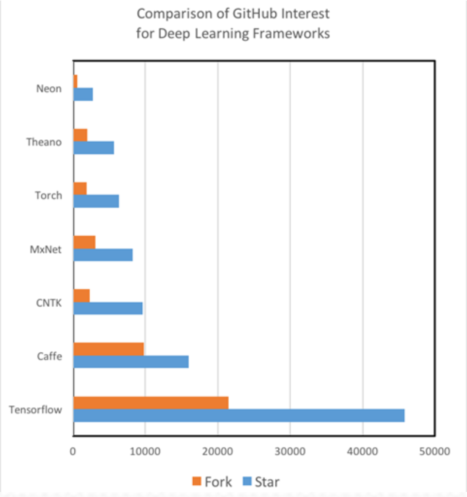
Inference (high performance inference)
TensorRT
MXNET
Caffe 2
Torch
Tensorflow

Nvidia GPU Direct

Benchmark
https://github.com/u39kun/deep-learning-benchmark

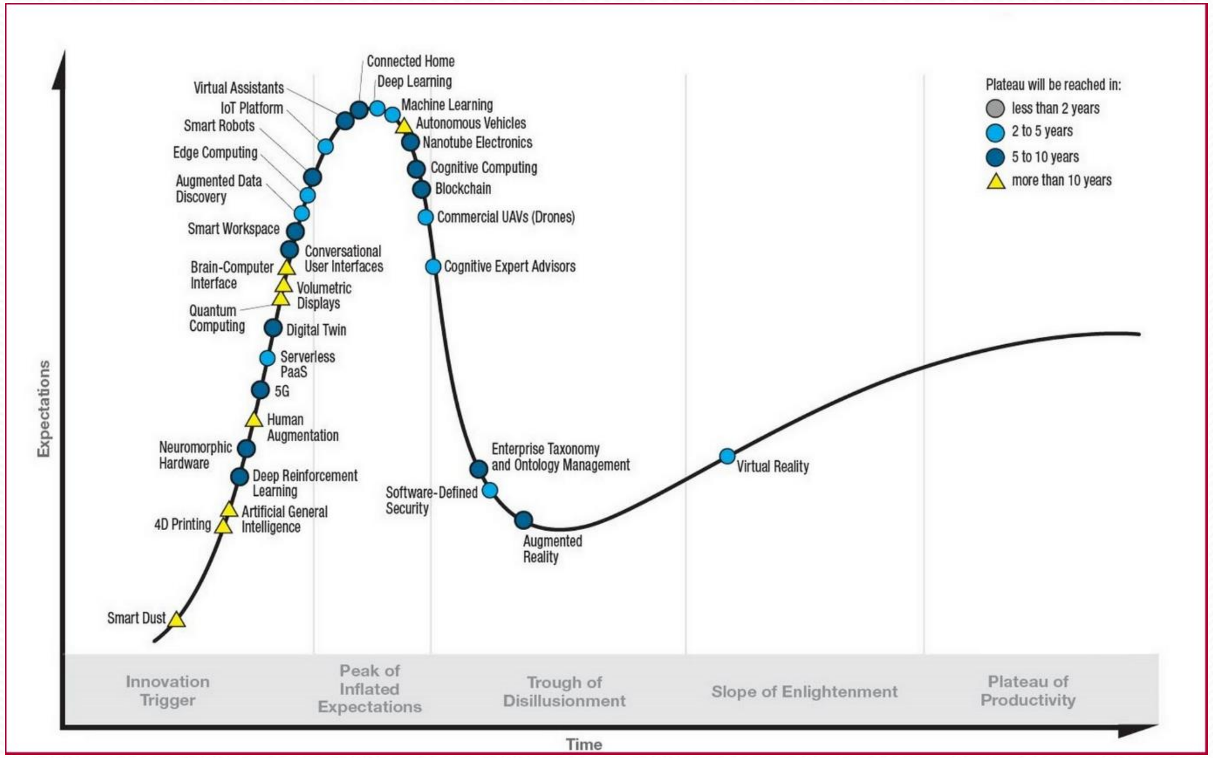






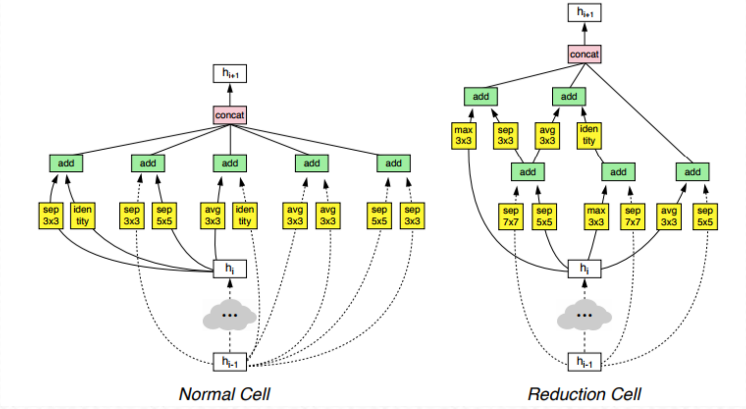
Is The Free Lunch Over?

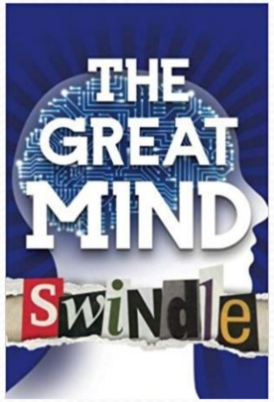
Кривая Гартнера – инновационные тренды



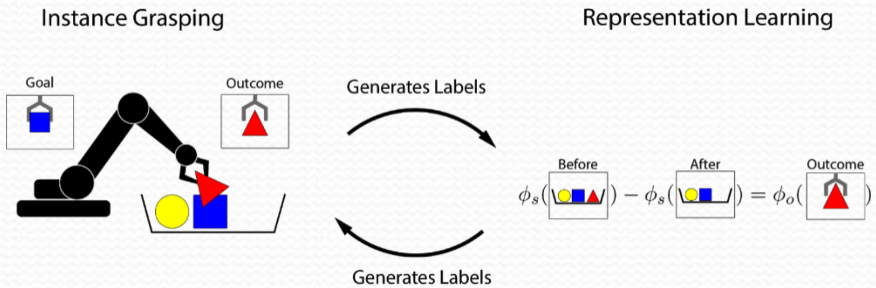
- Deep Learning на пике популярности, но есть проблемы – датасеты, интерпретация, часто требует сотен GPU, принцип обучения
- Reinforcement Learning – восходящий тренд, не столь требователен к ресурсам

Разные подходы к обучению требуют разной вычислительной мощности





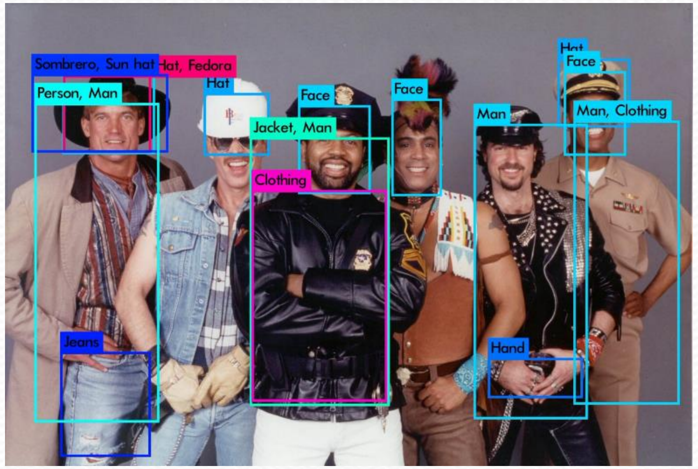
Learning Transferable Architectures for Scalable Image Recognition, 2018
512 GPUs



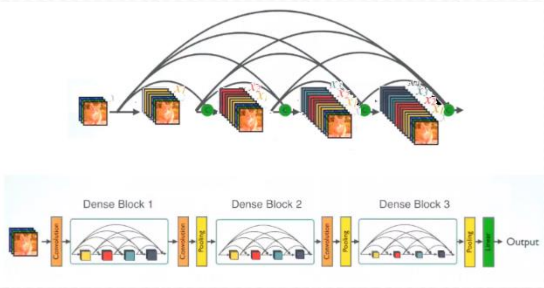
Grasp2Vec: Learning Object Representations from Self-Supervised Grasping – 1 GPU!!!

Позитивные тренды

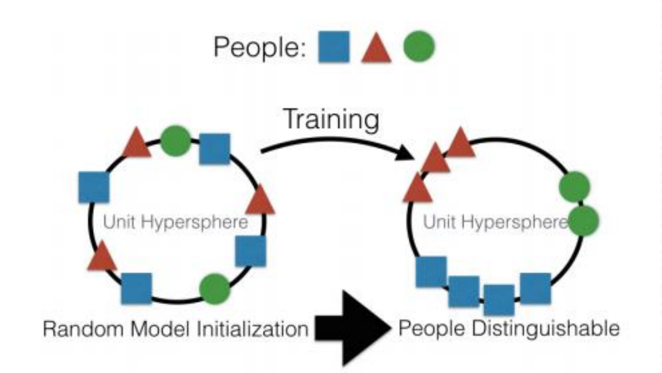
1. Reinforcement learning from Google Brain constructs NN, Learning Transferable Architectures for Scalable Image Recognition
2. Transfer learning
3. One shot learning
4. Network pruning
5. Mobile networks
6. Exotic networks
7. New datasets
8. New annotation tools!



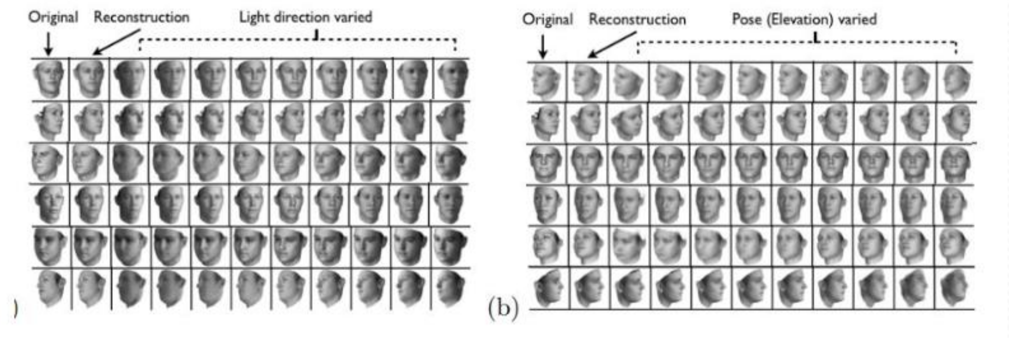
https://github.com/openimages/dataset



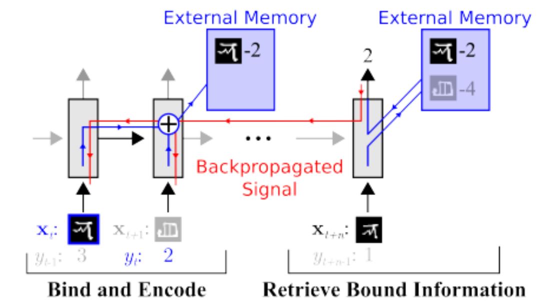
Exotic network



Агрегация по смыслу



Агрегация по параметрам



- Few Shot
- Memory

...

Ссылки

Stanford - https://cs231n.github.io/
Eugenio Culurciello - https://culurciello.github.io/
Русское сообщество - Slack OpenDataScience
Упражнения по DL - https://github.com/nehal96/Deep-Learning-ND-Exercises
Производительность железа: https://lambdalabs.com/blog/best-gpu-tensorflow-2080-ti-vs-v100-vs-titan-v-vs-1080-ti-benchmark/

Размышления на тему заката Deep Learning - https://habr.com/ru/company/recognitor/blog/455676/

Архитектуры 1 - https://habr.com/ru/company/wunderfund/blog/313696/
Архитектуры 2 - https://habr.com/ru/company/wunderfund/blog/313906/

Вычислительная фотография - https://vas3k.ru/blog/computational_photography/

Николенко С.И. и др. Глубокое обучение - https://www.ozon.ru/context/detail/id/142987816/

Комбинаторика и графы:
Hanjun Dai, et al., Learning Combinatorial Optimization Algorithms over Graphs, NIPS, 2017