Нейронные сети. Размер имеет значение

Дмитрий Николаев,
технический директор Smart Engines

Многообещающее начало I

В 1956 году Джон Маккарти, Марвин Минский, Клод Шеннон и Натаниэль Рочестер пишут заявку на двухмесячную конференцию по актуальным проблемам искусственного интеллекта:

Ограниченность современных компьютеров;
Использование машиной естественного языка;
Нейронные сети;
Теория сложности вычислений;
Самообучение;
Абстракция;
Случайность и творчество.

Джон Маккарти.
Ввел термин
“искусственный интеллект”,
разработал язык LISP.

Многообещающее начало II

*“Мы предлагаем исследование **искусственного интеллекта** сроком в **2 месяца** с участием **10 человек**… Мы считаем, что **существенное продвижение** в одной или более из этих проблем вполне возможно, если специально подобранная группа учёных будет работать над этим в течение лета.”*

Цена вопроса: $13500

Верю - не верю I

В 1943 году Уоррен Мак-Каллок и Уолтер Питтс в работе “A Logical Calculus of Ideas Immanent in Nervous Activity” предлагают математическую модель нейронных сетей, позволяющих вычислять логические выражения.

В 1960 году Фрэнк Розенблатт демонстрирует персептрон Марк-1 - нейросетевую машину, распознающую зрительные образы. Это вызвало всплеск энтузиазма.

В 1969 году Минский и Пейперт в книге “Perceptrons” показывают принципиальные ограничения однослойного персептрона и предполагают, что многослойные вряд ли лучше. Энтузиазм стремительно гаснет.

Верю - не верю II

В 1989 году Джордж Цыбенко доказывает, что нейронная сеть с одним скрытым слоем может аппроксимировать любую непрерывную функцию многих переменных с любой точностью. Воодушевляюще.

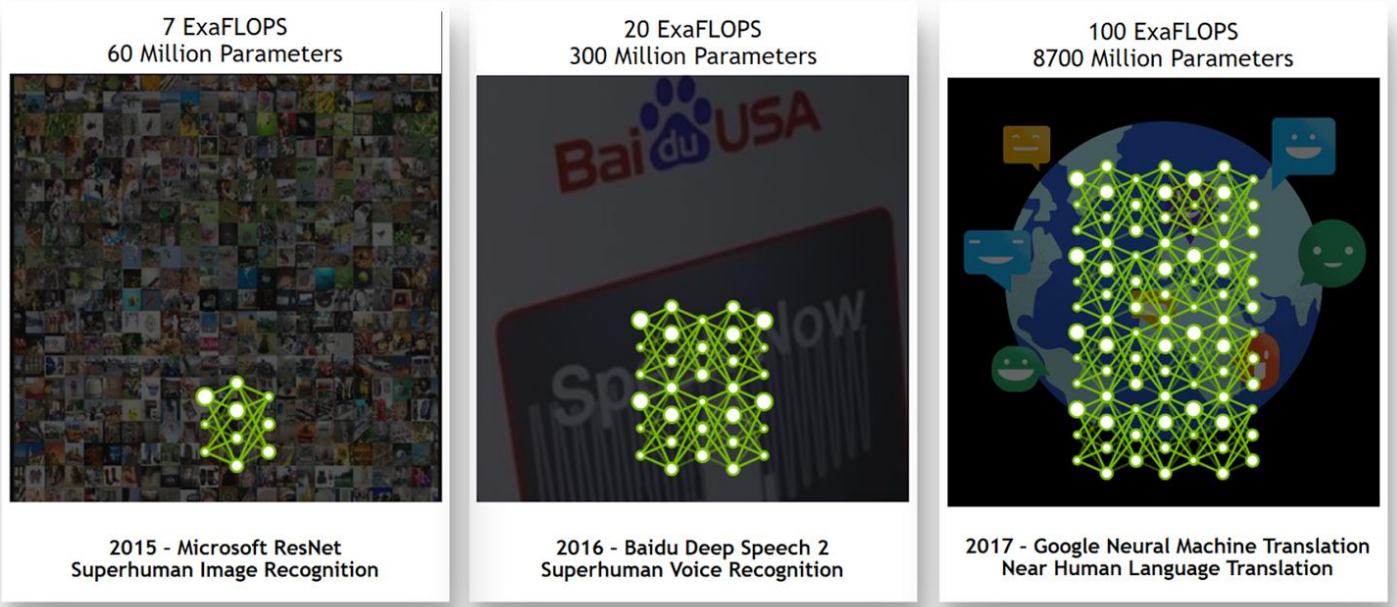
С 1985 по 2011 год доказывается множество теорем о том, что необходимое для решения многих задач число нейронов в подобных моделях может быть невероятно велико. Ужасно. Но не для глубоких сетей!

Искусственный интеллект уже на горизонте

Прошло 62 года. Актуальные проблемы искусственного интеллекта:

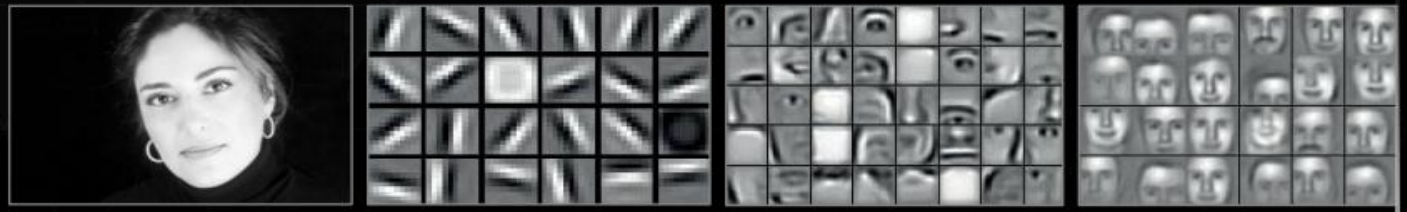
Ограниченность современных компьютеров;
Использование машиной естественного языка;
Глубокие нейронные сети;
Теория сложности вычислений;
Самообучение;
Абстракция;
Случайность и творчество.

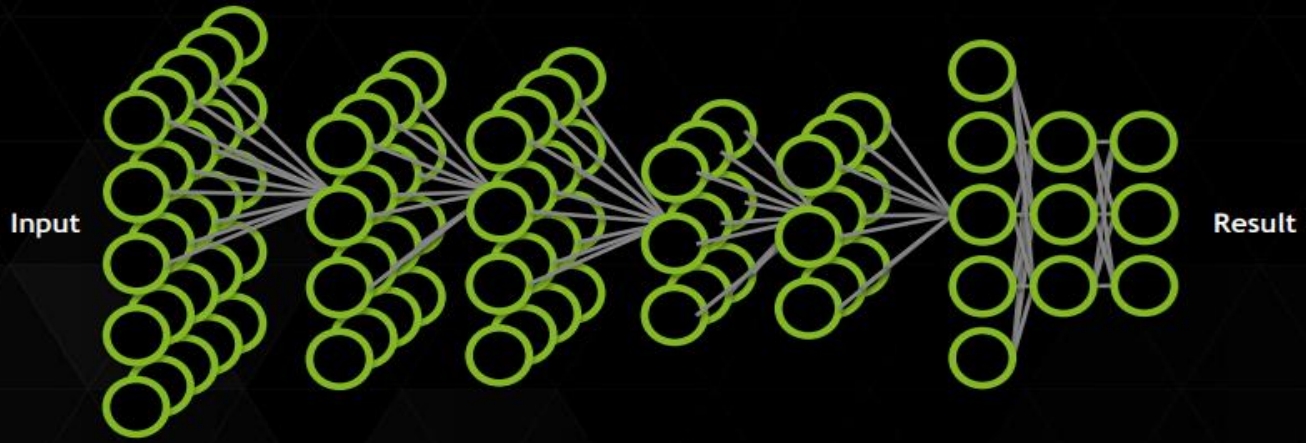
Прогресс ИИ глазами производителей «железа»



https://devblogs.nvidia.com/training-self-driving-vehicles-challenge-scale/

WHAT MAKES DEEP LEARNING DEEP?





Today's Largest Networks

~10 layers
1B parameters
10M images
~30 Exaflops
~30 GPU days

Human brain has trillions of parameters - only 1,000 more.

http://www.nvidia.com/content/events/geoInt2015/LBrown_DL.pdf

Глубже, больше, мощнее

Число слоев - 10.

Число параметров - 1 млрд.

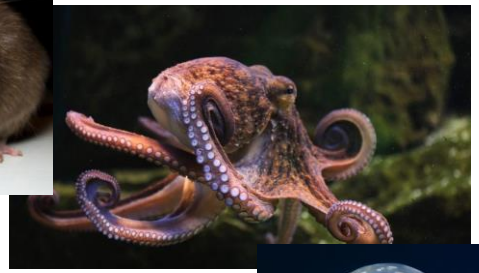
Число операций - не меньше, чем число параметров.

То есть они не только глубокие (много**слой**ные). У них много **параметров** и они требуют много **операций** для вычисления.

Больше вычислений - больше нейронов - больше ума - победа?

Количество нейронов в коре головного мозга

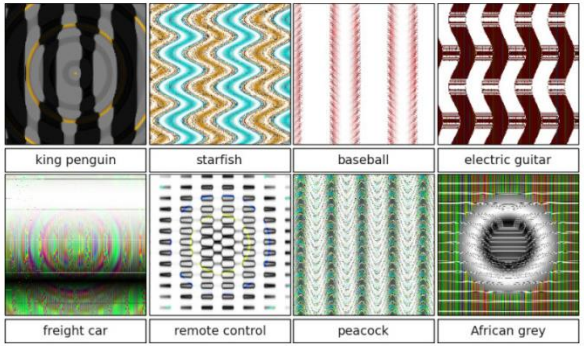
Вид	Количество нейронов
Шароголовый дельфин	**$3{,}7 * 10^{10}$**
Человек	$1{,}6 * 10^{10}$
Шимпанзе	$6{,}2 * 10^{8}$
Мышь	$1{,}4 * 10^{7}$
Осьминог	$3{,}0 * 10^{6}$
Таракан	$1{,}0 * 10^{6}$
Фруктовая муха	$2{,}5 * 10^{5}$
Медуза	800



Сложнее, точнее, быстрее

Большой объем вычислений - не то, чем можно гордиться. При прочих равных это - чистые потери.

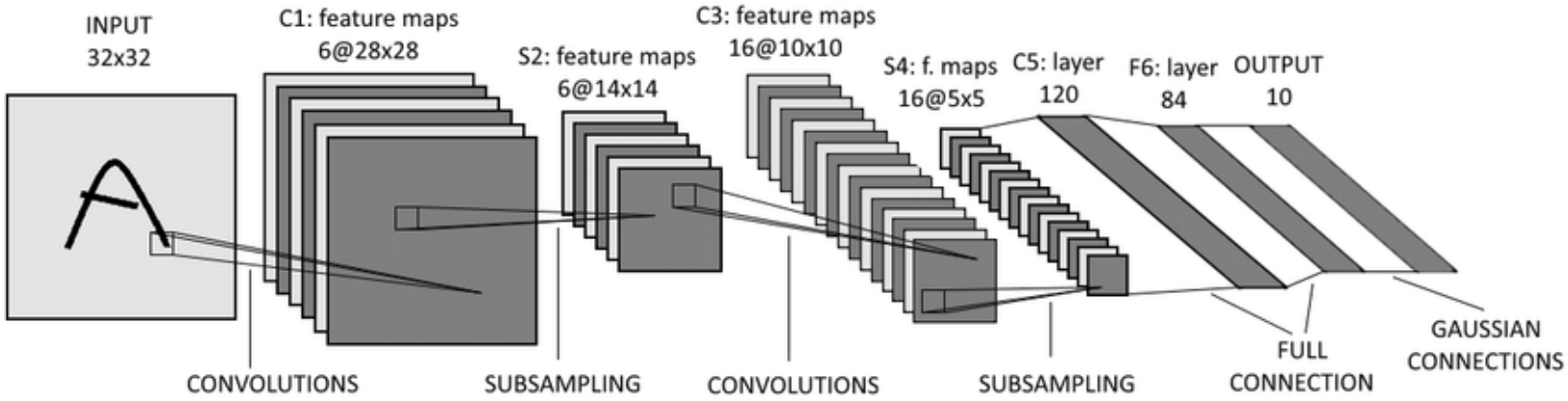
Большое число параметров - это риск дурацких.



Даже если правда, что иначе задачу не решить, это - необходимое зло.

Важна **сложность**, но **задачи**, а не сети. Важна **точность** решения этой задачи. Важно **быстродействие**.

Число операций: встретились параметры с мегапикселями



LeNet-5, 7 слоев, ~60k параметров

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998.

Высокая детальность входных данных - загрязнение окружающей среды.

Специализированные устройства для нейросетевых вычислений

2017, Google Tensorflow Processing Unit:

ASIC для быстрой обработки нейронных сетей (15–30X большая производительность, и 30–80X большая производительность на ватт, чем у современных CPU и GPU) https://cloud.google.com/blog/big-data/2017/05/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu

2017, Microsoft Project Brainwave:

обучение в облаке на базе FPGA

Нейроморфные чипы

Отличаются высокой энергоэффективностью за счет импульсных нейронов, имеют эффективную архитектуру для параллельного доступа в память.

- 2014, DARPA SyNAPSE, TruNorth
- 2015, SpiNNaker Project:
  планируется объединить до миллиона процессоров ARM для моделирования человеческого мозга
- 2017, Intel Loihi
  асинхронные импульсные нейроны, 130 000 нейронов и 130 миллионов синапсов, возможность самообучения
- Пока не достигают точности глубоких нейронных сетей

Попроще бы как-нибудь умножать...

Yann LeCun: “Getting excellent results on ImageNet is easily achieved with a convolutional net with something like 8- to 16-bit precision on the neuron states and weights.”

2011, Google, 8-битная квантизация нейронных сетей [1]

2015, IBM, обучение в 16-битных целых числах [2]

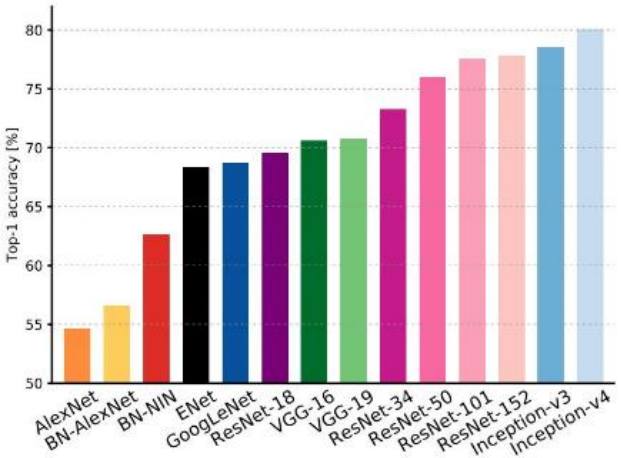
2016, Google, квантованные целочисленные нейронные сети в tensorflow

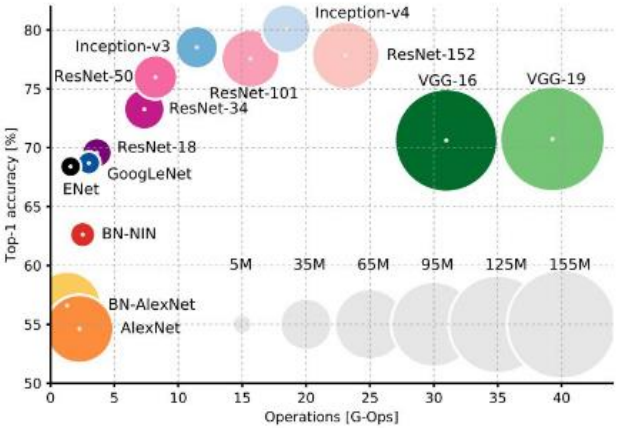
2017, SmartEngines, послойное дообучение 8-битных сетей [3]

2018, SmartEngines, глубокие алгебраические сети без умножений

1. Vanhoucke, V. et al. "Improving the speed of neural networks on CPUs." Proc. Deep Learning and Unsupervised Feature Learning NIPS Workshop. Vol. 1. 2011.
2. Gupta, S. et al. "Deep learning with limited numerical precision." International Conference on Machine Learning. 2015.
3. Ilin, D. et al. "Fast integer approximations in convolutional neural networks using layer-by-layer training", Proc. SPIE 10341, Ninth International Conference on Machine Vision (ICMV 2016), 103410Q (17 March 2017)

Вычислительная сложность и качество распознавания глубоких архитектур на ILSVRC

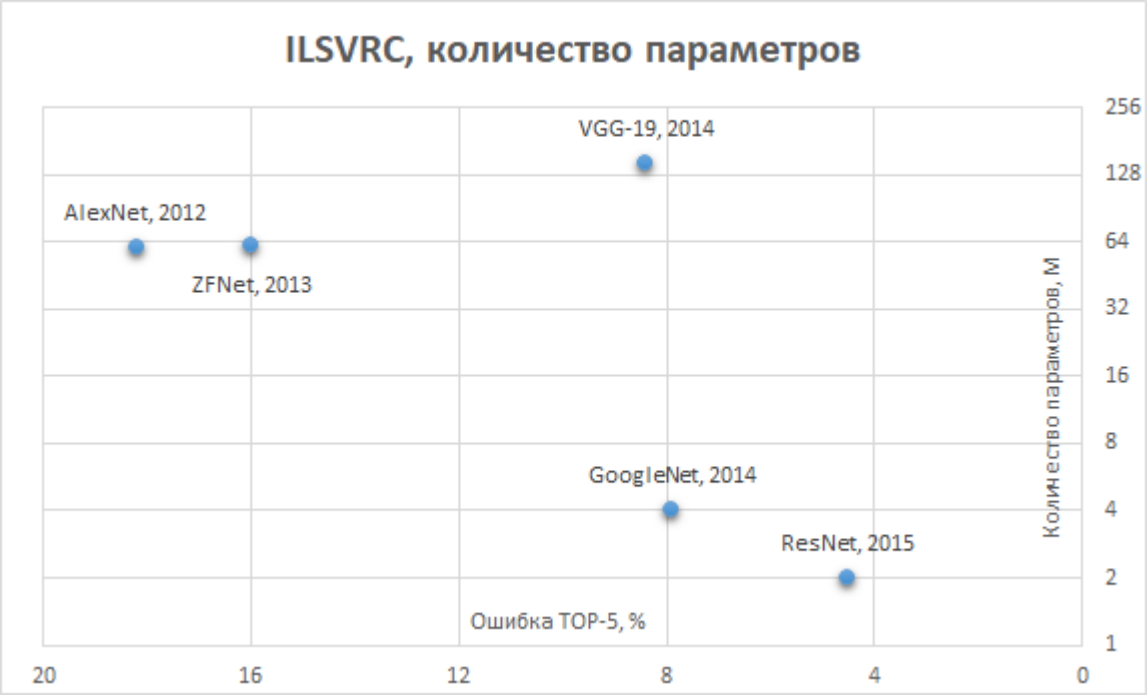




An Analysis of Deep Neural Network Models for Practical Applications, 2017.

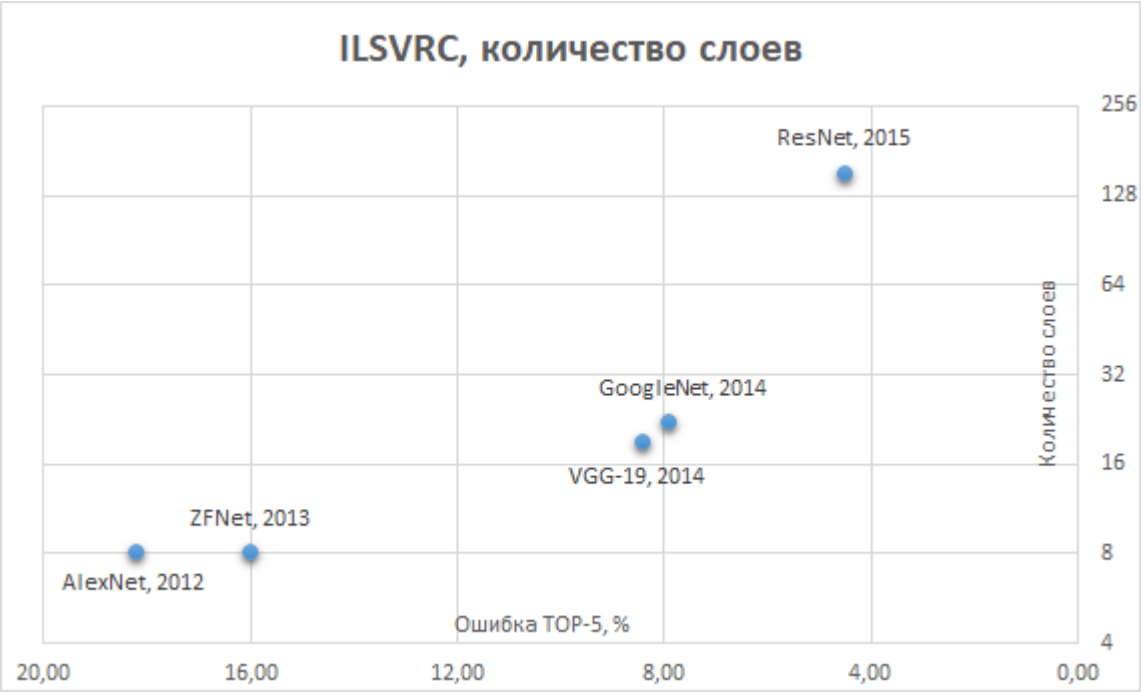
ILSVRC, количество параметров сетей

	TOP-5	Кол-во весов
AlexNet, 2012	18,2	60
ZFNet, 2013	16,0	62
VGG-19, 2014	8,4	144
GoogleNet, 2014	7,9	4
ResNet-152, 2015	4,5	2



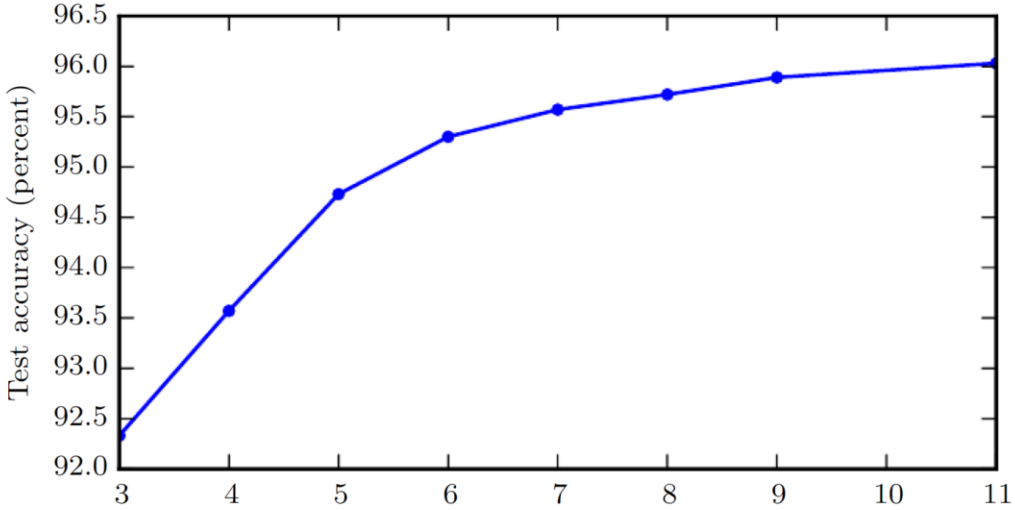
ILSVRC, количество слоев

	TOP-5	Кол-во слоев
AlexNet, 2012	18,2	8
ZFNet, 2013	16,0	8
VGG-19, 2014	8,4	19
GoogleNet, 2014	7,9	22
ResNet-152, 2015	4,5	152



Влияние количества слоев I

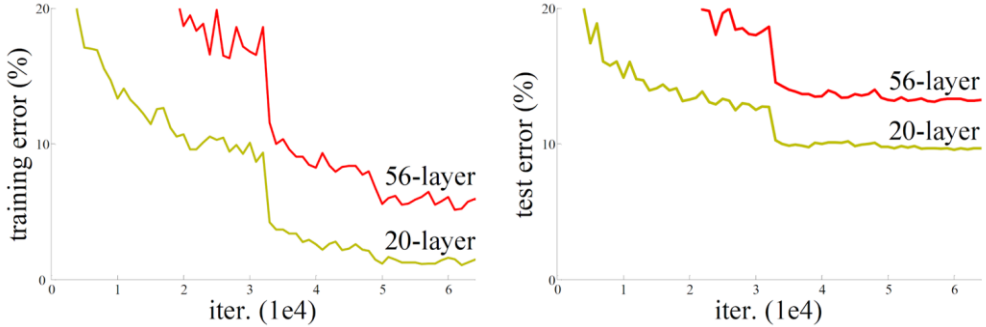
Более глубокие сети показывают лучшую способность к обобщению. Но это работает только до какого-то момента



[1] I. Goodfellow et al., **Multi-digit number recognition from Street View imagery using deep convolutional neural networks**, ICLR 2014

Влияние количества слоев II

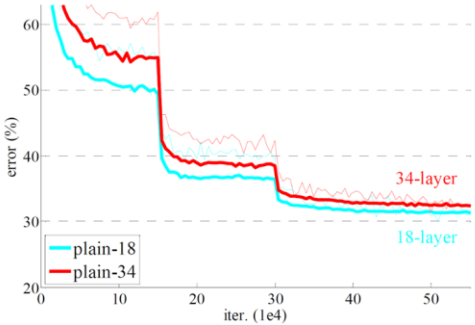
Дальнейшее увеличение количества слоев *плоской* сети приводит к деградации качества, как на обучающем, так и на тестовом наборе

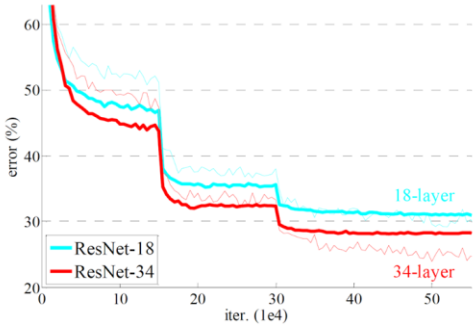


[2] K. He et al., **Deep Residual Learning for Image Recognition**, CVPR 2016

Влияние количества слоев: ResNet

Иное поведение демонстрируют residual networks, обычно содержащих огромное количество слоев

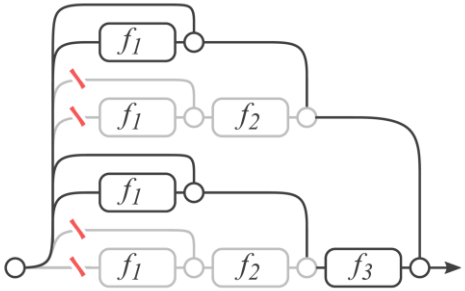




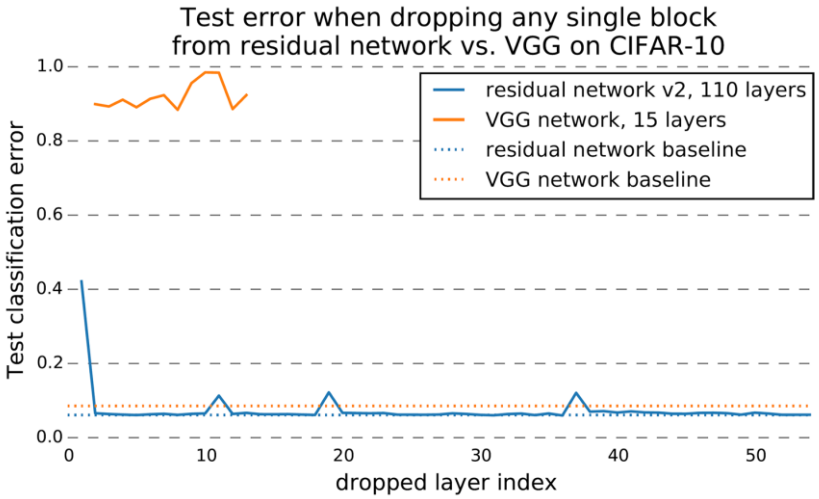
[2] K. He et al., **Deep Residual Learning for Image Recognition**, CVPR 2016

ResNet как ансамбль моделей

Но удаление слоев из ResNet приводит лишь к незначительной потере в качестве



(a) Deleting f_2 from unraveled view

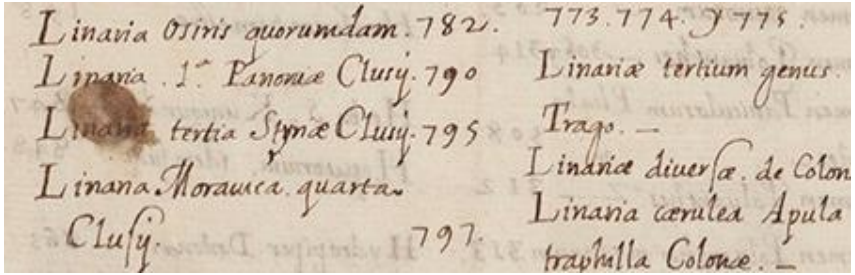


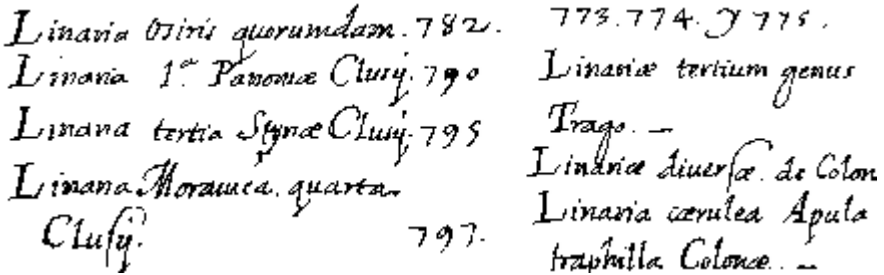
[3] A. Veit, **Residual networks behave like ensembles of relatively shallow networks**, NIPS 2016

Document Image Binarization Contest (DIBCO’17)

Прошедший конкурс: https://vc.ee.duth.gr/dibco2017

Не за горами уже следующий: https://vc.ee.duth.gr/h-dibco2018/





DIBCO’17: Результаты

IAPR ICDAR 2017

IAPR International Conference on Document Analysis and Recognition 2017

ICDAR2017 Competition on
Document Image Binarization (DIBCO 2017)

... and the winner is ...

Dmitrii Ilin, Pavel Bezmaternykh, Dmitry Nikolaev
Smart Engines Ltd, Moscow, Russia

Ivan-Marc Ogier, Competition Chair

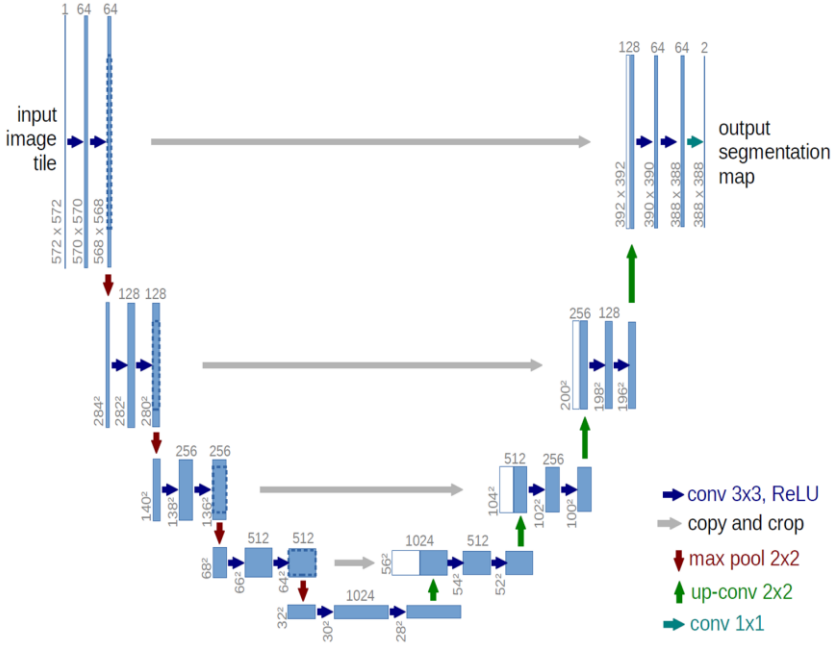
Marcus Liwicki, Competition Chair

I. PRATIKAKIS

Лучший метод, не использующий нейронные сети, занял только седьмое место.

№	Метод	Результат
1	**U-net архитектура** (Smart Engines)	309
2	FCN (VGG подобная архитектура) + постфильтрация	455
3	Ансамбль из 3-х DSN, с 3-х уровневым выходом, работающие на патчах разных масштабов	481
4	Ансамбль из 5-ти FCN вход: патчи разных масштабов + бинаризованные методом Howe + RD признаки	529
...	...	...
7	Морфология + метод Howe + постобработка	635

Архитектура U-Net



Большой класс известных алгоритмов бинаризации явно выразим в подобной архитектуре.

Используемые слои:

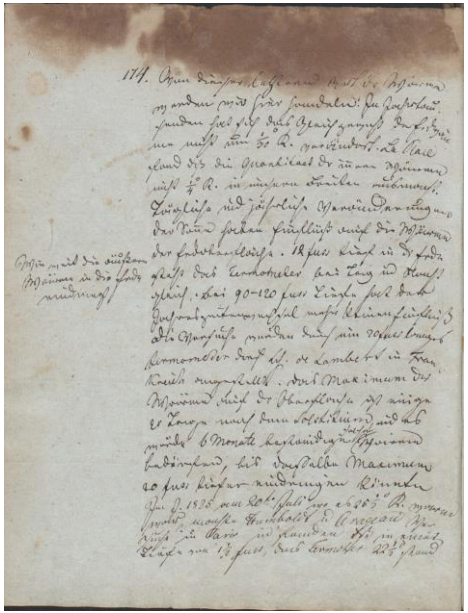
Convolution(16, (3, 3)), MaxPooling(2, 2)
Convolution(32, (3, 3)), MaxPooling(2, 2)
Convolution(64, (3, 3))
UpSampling(2, 2), Convolution(32, (3, 3))
UpSampling(2, 2), Convolution(16, (3, 3))

Примерно ~15.3 миллиона операций на изображении размером 128x128

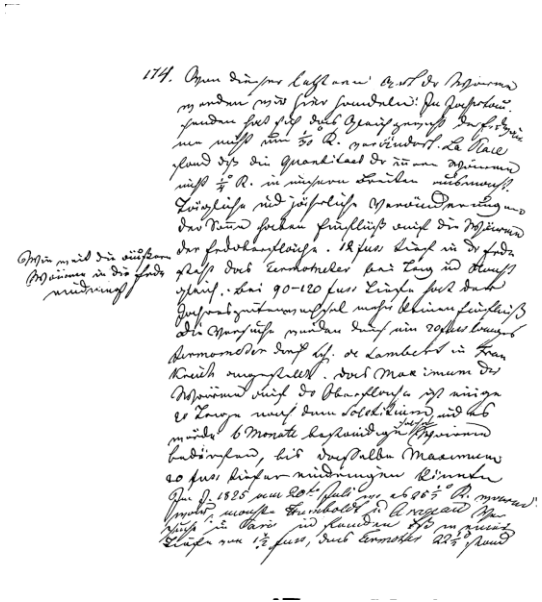
Число параметров ~52 000

https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net

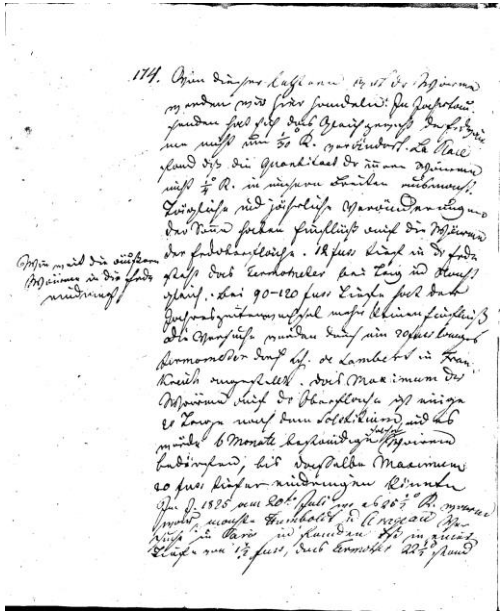
Оно того стоило?



Исходное изображение



Результат работы сети



Результат работы метода Ниблэка (2 параметра !!!)

В вашем смартфоне, на CPU

Сеть для распознавания номеров кредитных карт:

3 слоя, ~12K параметров

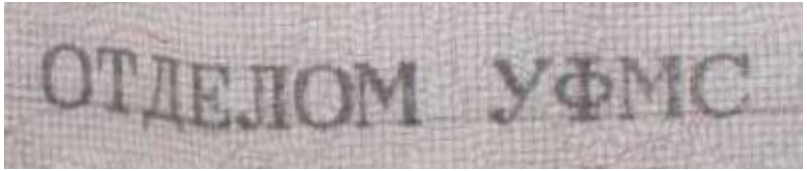
Сеть для распознавания текста в паспорте РФ:

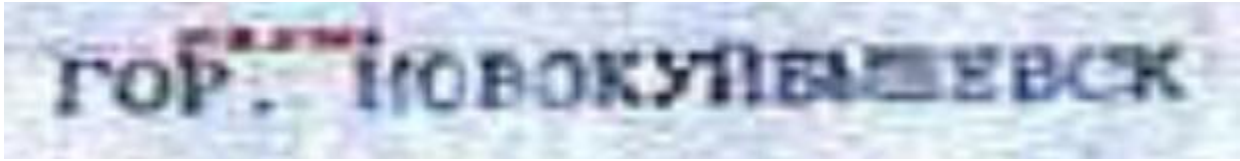
5 слоев, ~11K параметров

Задача сегментации считается более сложной

10 слоев, ~3,5K параметров

Примеры распознаваемых строк из паспорта РФ





Итого

За 60 лет пути технологии искусственного интеллекта фантастически усовершенствовались. Но ни одна из проблем, объявленных на старте важными, так толком и не решена.

Не ищите специалистов по искусственному интеллекту. Ищите специалистов по решению ваших задач.

Спасибо!