

Основы практического использования нейронных сетей.

Лекция 7. Обзор современных глубоких сетей для
классификации изображений.

Дмитрий Буряк.
к.ф.-м.н.
dyb04@yandex.ru

Задача классификации изображений

- ❑ Множество изображений.
- ❑ Каждое изображение содержит один объект какого-то класса.
- ❑ Классы объектов заданы.
- ❑ Необходимо определить к какому из известных классов относится изображенный объект.
- ❑ Не требуется определять позицию, где находится объект.
- ❑ Два варианта задачи классификации:
 - Каждое изображение содержит только объекты заданных классов;
 - Есть изображения объектов не из заданного набора классов.
- ❑ Точность алгоритма классификации: $\frac{M}{N} * 100\%$,
M – число изображений с правильно определенным классом,
N – общее количество изображений.
- ❑ Пример: классификация изображений MNIST

label = 5



label = 0



label = 4



label = 1



label = 9



База изображений ImageNet

❑ ImageNet

- Создана в 2009
- Более 14М изображений с аннотациями
- 1М изображений с отмеченными объектами
- Более 20К классов (несколько сотен изображений каждого класса)



❑ ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

- Проводился с 2010г по 2017г;
- Классификация изображений, обнаружение объектов.



❑ ILSVRC. Классификация изображений

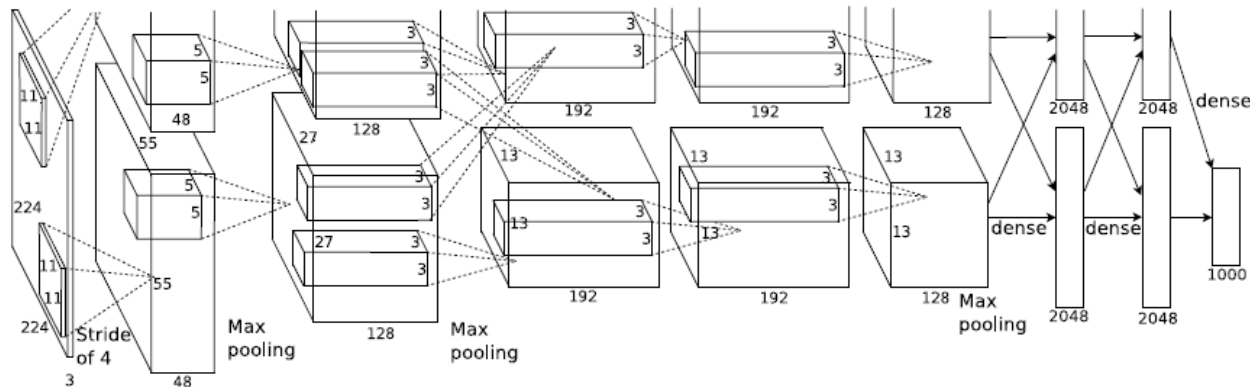
- 1.2М изображений (обучающая выборка)
- 100К тестовых изображений
- 1000 классов



Победитель ILSVRC-2012. AlexNet

- ❑ AlexNet (A. Krizhevsky et al, 2012)

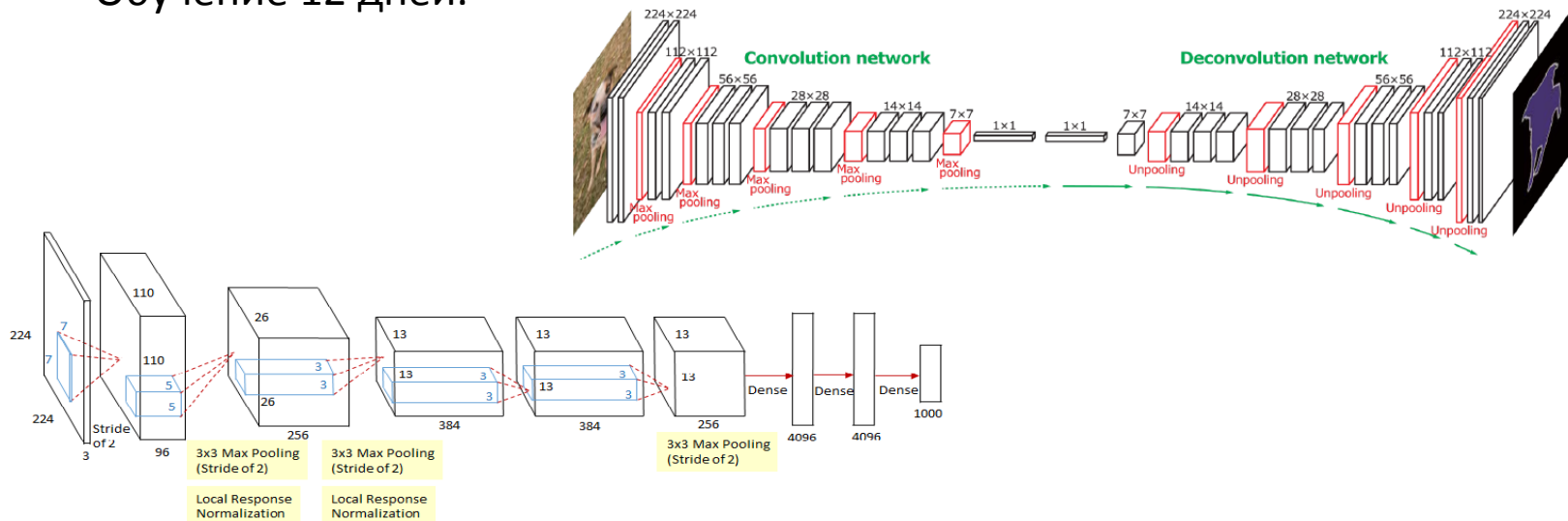
- Первая сверточная сеть победитель ILSVRC;
- Ошибка Top-5: **15.3%** (второй результат SIFT – 26.2%);
- Сверточная НС, 8 слоев (5 сверточных), 60М параметров;
- ReLU, Dropout, Увеличение обучающей выборки (сдвиг, отражение);
- Обучение заняло 5 дней на нескольких GPU.



Победитель ILSVRC-2013. Clarifai

□ ZFNet (Clarifai – стартап авторов ZFNet) – (М. Zeiler et al, 2014)

- Ошибка Top-5: **11.2%** (несколько сетей);
- Визуализация активаций внутренних слоев - Deconvnet;
- Анализ внутренних шаблонов → тонкая настройка AlexNet;
- Фильтры первого слоя (11x11) → (7x7); шаг свертки первого слоя 4 → 2;
- Обучение 12 дней.



VGG Net

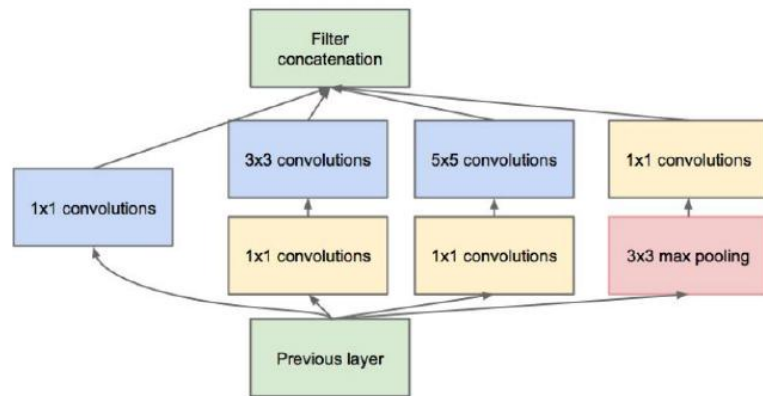
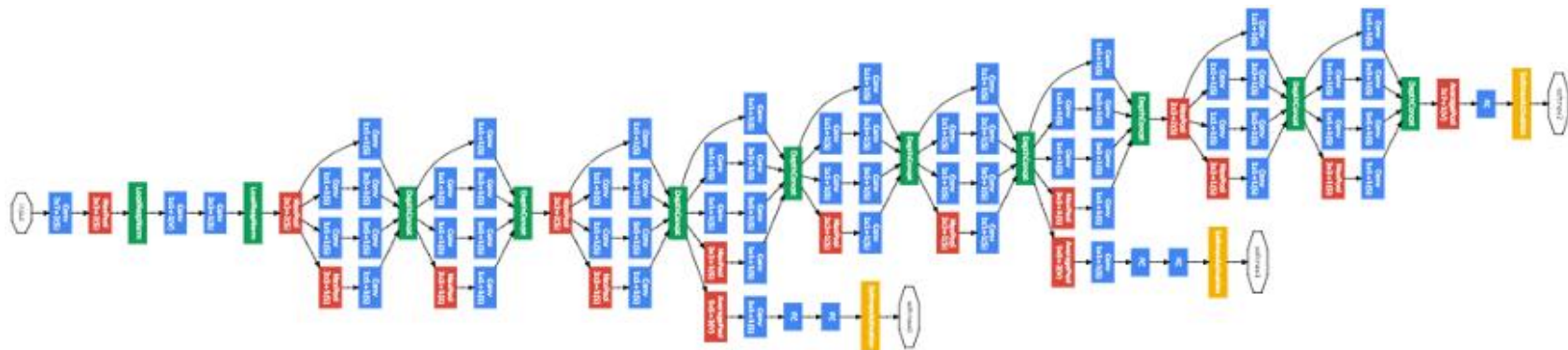
- VGG (K.Simonyan, A. Zisserman, 2014)
 - Ошибка Top-5: **7.3%** (ILSVRC-2014);
 - Классическая CNN;
 - Применение ядер 3x3;
 - $(3 \times 3) + (3 \times 3) \sim (5 \times 5)$; $(3 \times 3) + (3 \times 3) + (3 \times 3) \sim (7 \times 7)$;
 - Обучение 4GPUx(3 недели);
 - Активно применяется при решении задач.

D
16 weight layers
)
conv3-64
conv3-64
conv3-128
conv3-128
conv3-256
conv3-256
conv3-256
conv3-512
conv3-512
conv3-512
conv3-512
conv3-512
conv3-512
maxpool
FC-4096
FC-4096
FC-1000
soft-max

Победитель ILSVRC-2014. GoogLeNet

□ GoogLeNet – (K. Szegedy et al, 2015)

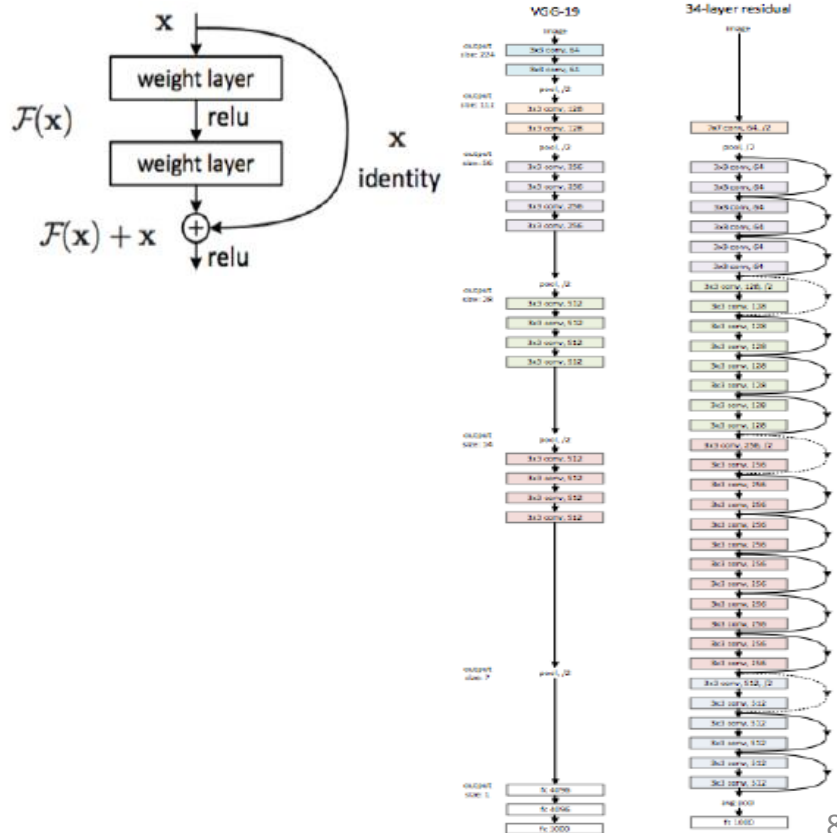
- Ошибка Top-5: **6.7%**;
- «Параллельные свертки-Конкатенация»:
Inception модуль.
- 22 модуля, более 50 сверточных слоев;
- Несколько выходных слоев;
- Размер 55Мб (VGG16 – 490Мб);
- Обучение 7 дней на нескольких GPU;



Победитель ILSVRC-2015. ResNet

□ ResNet – (K. He et al, 2015)

- Ошибка Top-5: **3.57%**;
- Рост глубины $\rightarrow$ меньше ошибка $\rightarrow$ сложность оптимизации;
- «Остаточные» связи – единичное отображение, нет дополнительных параметров;
- «Остаточные» соединения $\rightarrow$ лучше распространение градиента
- 152 слоя, ядра 3x3;
- Обучение 21 день на 8 GPU.

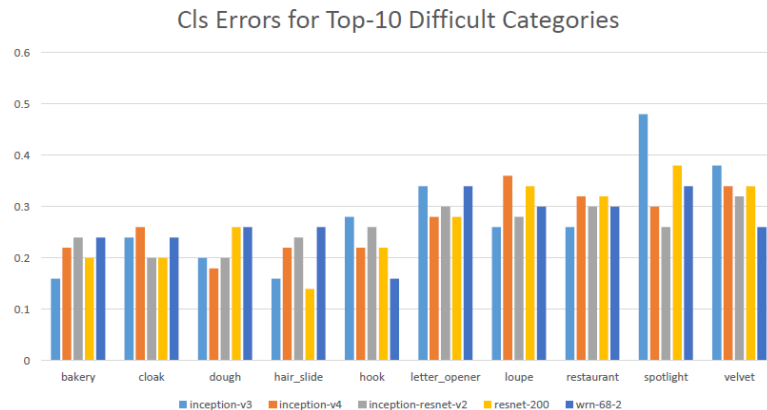


Победитель ILSVRC-2016. Trimps-Soushen

❑ Trimps-Soushen

- Ошибка Top-5: **2.99%**;
- Использование предобученных сетей для выявления Top-10 наиболее сложных категорий в ImageNet;
- Inception (v3, v4), Inception-ResNet-v2, Pre-Activation ResNet-200, Wide ResNet;
- Нет лучшей сети для всех категорий из Top-10 + слабая коррелированность решений → объединение решений.

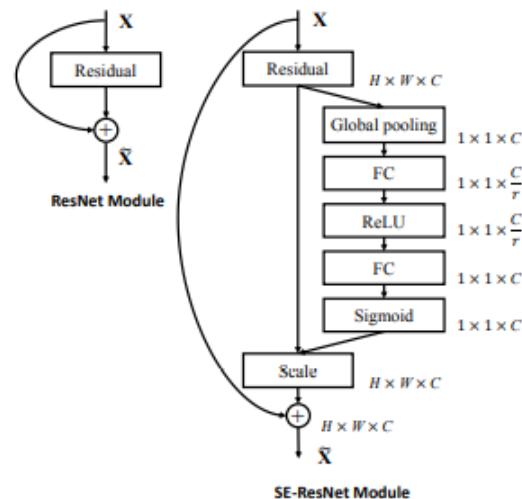
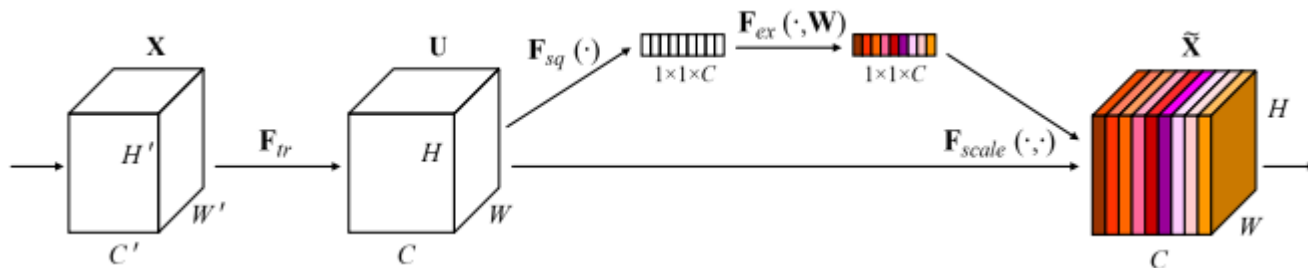
	Inception-v3	Inception-v4	Inception-Resnet-v2	Resnet-200	Wrn-68-3	Fusion (Val.)	Fusion (Test)
Err. (%)	4.20	4.01	3.52	4.26	4.65	2.92 (-0.6)	2.99



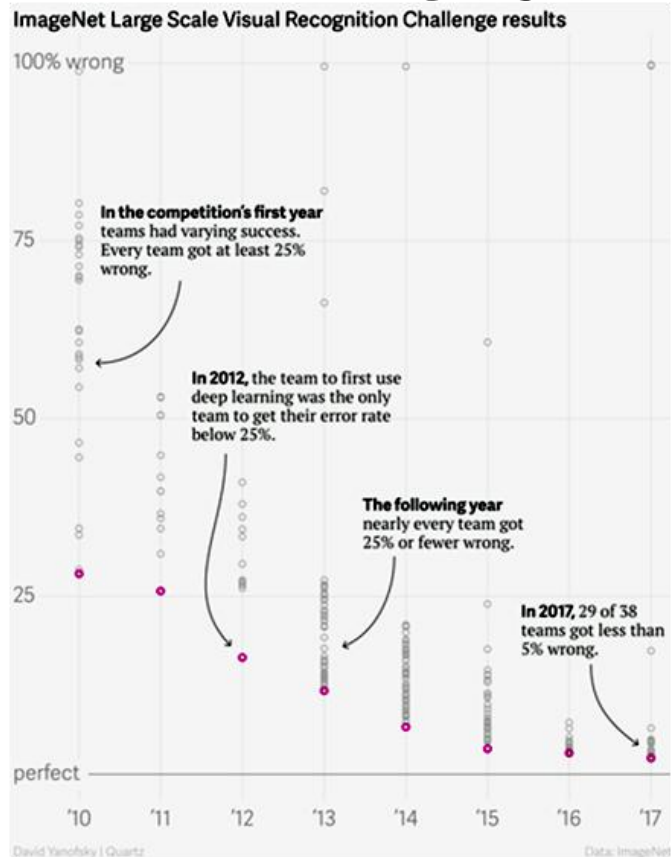
Победитель ILSVRC-2017. SENet

□ SENet-154 – (J. Hu et al, 2017)

- Ошибка Top-5: **2.25%**;
- Взвешивание карт признаков;
- «Squeeze and Excitation» (SE) модуль;
- Интеграция в различные существующие архитектуры: Inception, ResNet и т.д.
- Победитель SE+ResNeXt;
- Обучение на 64 GPU для поддержки пакетов размером 2048 изображений.

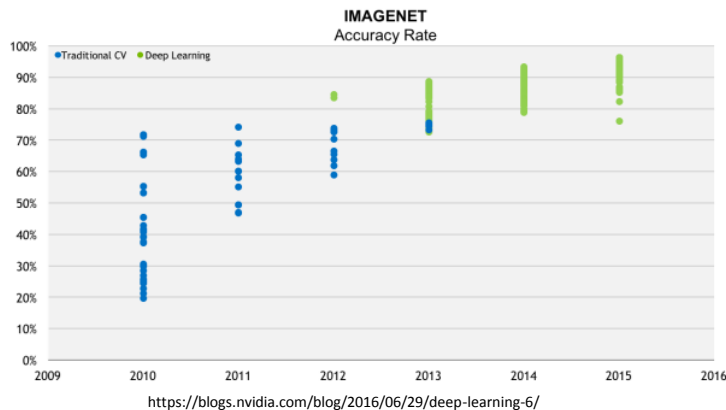


Статистика ILSVRC

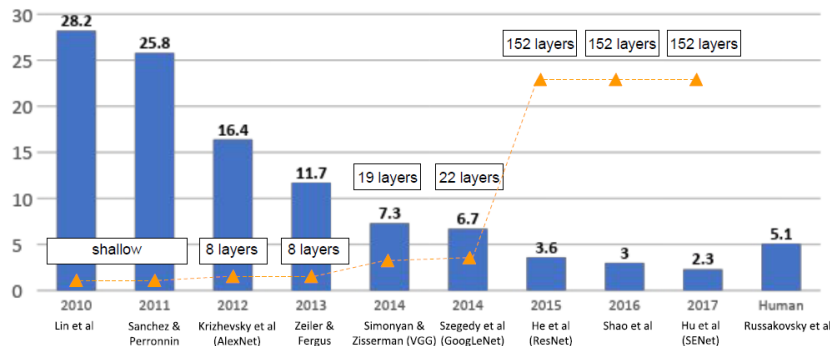


Сеть	Ошибка Top-5
AlexNet	15.3%
ZFNet	11.2%
VGG	7.3%
GoogLeNet	6.7%
ResNet	3.57%
Trimps-Soushen	2.99%
SENet-154	2.25%

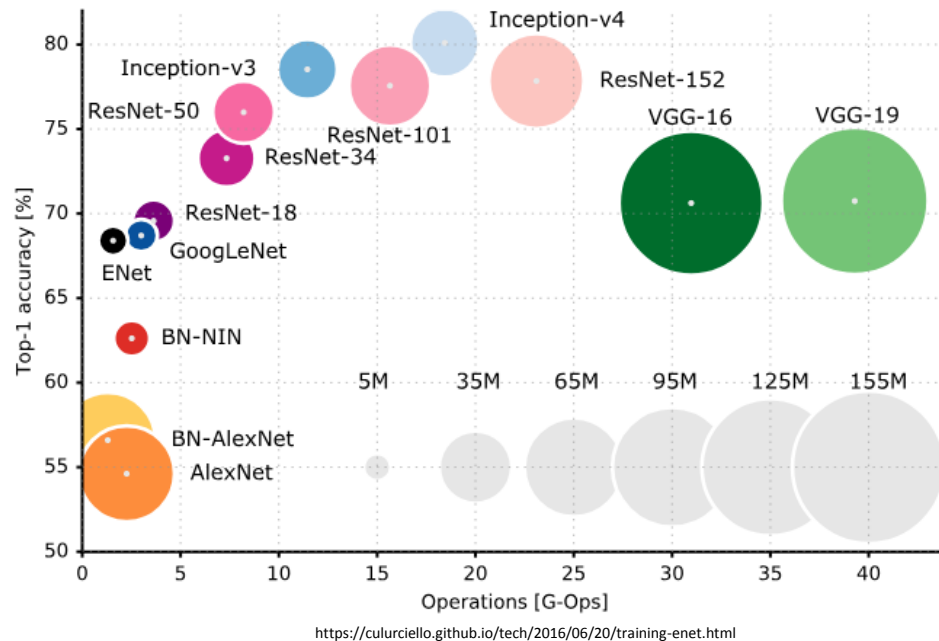
Развитие НС для ILSVRC



ImageNet Large Scale Visual Recognition Challenge (ILSVRC) winners



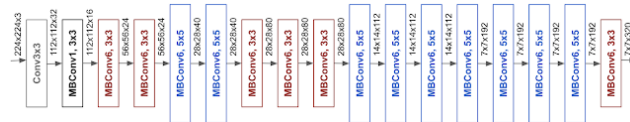
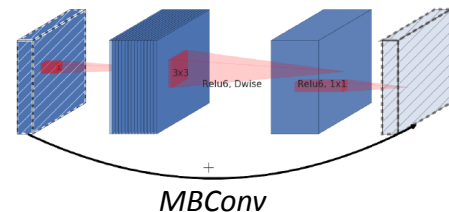
http://cs231n.stanford.edu/slides/2018/cs231n_2018_lecture09.pdf



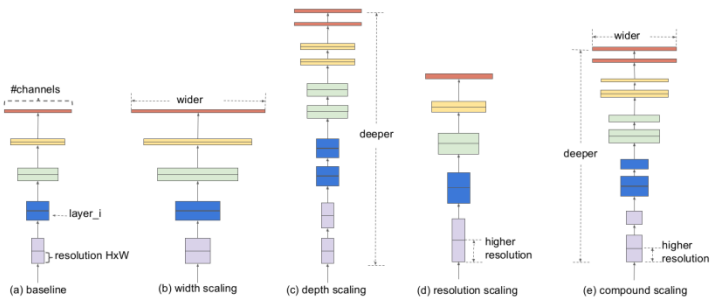
2019. EfficientNet

□ EfficientNet (M. Tan and Q. Lee et al, 2019)

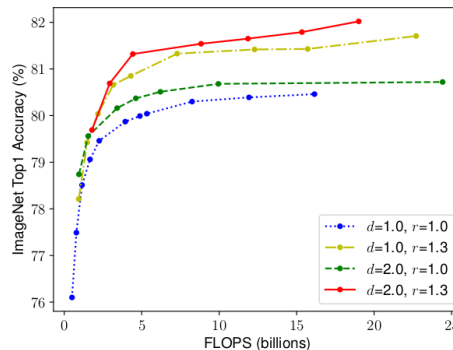
- Ошибка Top-1: 15.7%;
- Блоки MBConv
- NAS → Базовая модель EfficientNet-B0 →
→ Смешанное масштабирование архитектуры



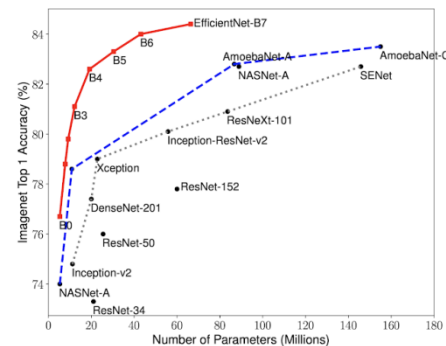
Архитектура EfficientNet-B0



Варианты масштабирования архитектуры



Масштабирование «ширины» сети для различных архитектур



2021. NFNet

❑ NFNet – Normalizer-Free Network – (A. Brock et al, 2021)

- Ошибка Top-1: 13.5%;
- Без Batch Normalization;
- Масштабирование весов;

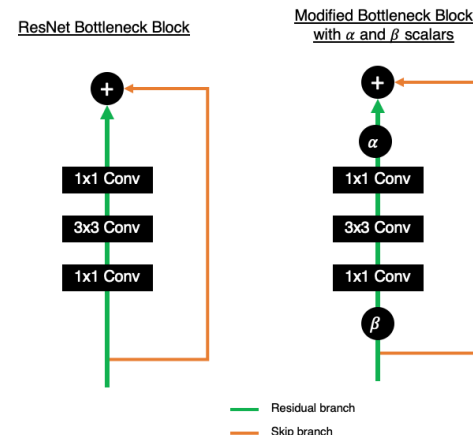
$$\hat{W}_{ij} = \frac{W_{ij} - \mu_i}{\sqrt{N}\sigma_i}$$

μ_i σ_i – среднее и среднеквадратичное отклонением весов

- Контроль значений градиентов (Gradient clipping).

$$G_i^\ell \rightarrow \begin{cases} \lambda \frac{\max(\|W_i^\ell\|, \epsilon)}{\|G_i^\ell\|} G_i^\ell & \text{if } \frac{\|G_i^\ell\|}{\|W_i^\ell\|} > \lambda \\ G_i^\ell & \text{otherwise} \end{cases}$$

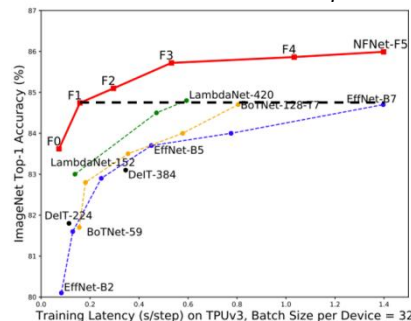
- Оптимизация архитектуры SE+ResNeXt.



$$\alpha = 0.2 \quad \beta_i = \sqrt{\text{Var}(h_i)}$$

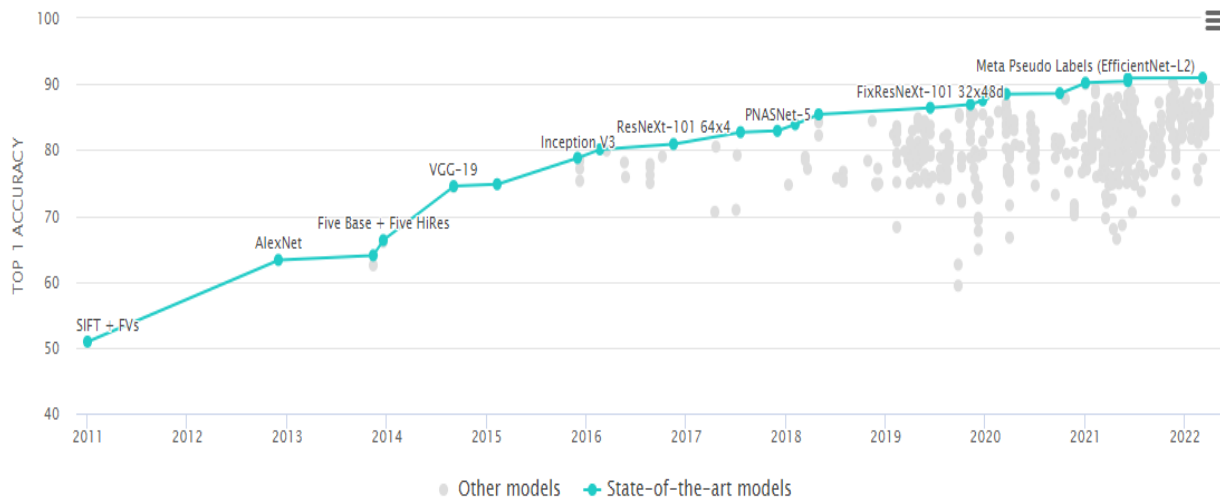
$$\text{Var}(h_i) = \text{Var}(h_{i-1}) + \alpha^2$$

h_i - веса



Современное состояние

❑ Ошибка Top-1



Rank	Model	Top 1 Accuracy	Top 3 Accuracy	Number of params	Extra Training Data	Paper	Code	Result	Year	Tags
1	Model soups (ViT-G/14)	90.94%		1843M	✓	Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time		📄	2022	Transformer, JFT-3B
2	CoAtNet-7	90.88%		2440M	✓	CoAtNet: Marrying Convolution and Attention for All Data Sizes		📄	2021	Conv + Transformer, JFT-3B
3	ViT-G/14	90.45%		1843M	✓	Scaling Vision Transformers		📄	2021	Transformer, JFT-3B
4	CoAtNet-6	90.45%		1470M	✓	CoAtNet: Marrying Convolution and Attention for All Data Sizes		📄	2021	Conv + Transformer, JFT-3B
5	Meta Pseudo Labels (EfficientNet-L2)	90.2%	98.8%	480M	✓	Meta Pseudo Labels		📄	2021	EfficientNet, JFT-3B
6	SwinV2-G	90.17%			✓	Swin Transformer V2: Scaling Up Capacity and Resolution		📄	2021	Transformer
7	Florence-Coswin-H	90.05%	99.02%		✓	Florence: A New Foundation Model for Foundation Models		📄	2021	Transformer

❑ Большинство эффективных НС основаны на трансформерах

Ссылки

- ❑ Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.
- ❑ Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *European Conference on Computer Vision*. Springer International Publishing, 2014.
- ❑ Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for largescale image recognition." *arXiv preprint arXiv:1409.1556*(2014).
- ❑ Szegedy, Christian, et al. "Going deeper with convolutions." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
- ❑ He, Kaiming, et al. "Deep residual learning for image recognition." *arXiv preprint arXiv:1512.03385* (2015).
- ❑ Hu, Jie, et al. "Squeeze-and-Excitation Networks." *arXiv preprint arXiv:1709.01507* (2017).
- ❑ Mingxing Tan, Quoc V. Le. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*, 2019.
- ❑ Andrew Brock, Soham De, Samuel L. Smith, Karen Simonyan. *High-Performance Large-Scale Image Recognition Without Normalization*, 2021

Вопросы

- ☐ В чем отличие базовых блоков сети GoogLeNet и ResNet?
- ☐ Какое назначение SE модуля?
- ☐ Какая идея лежит в основе построения семейства сетей EfficientNet?