

# Современные методы распознавания лиц

Петюшко А. А., к.ф.-м.н.

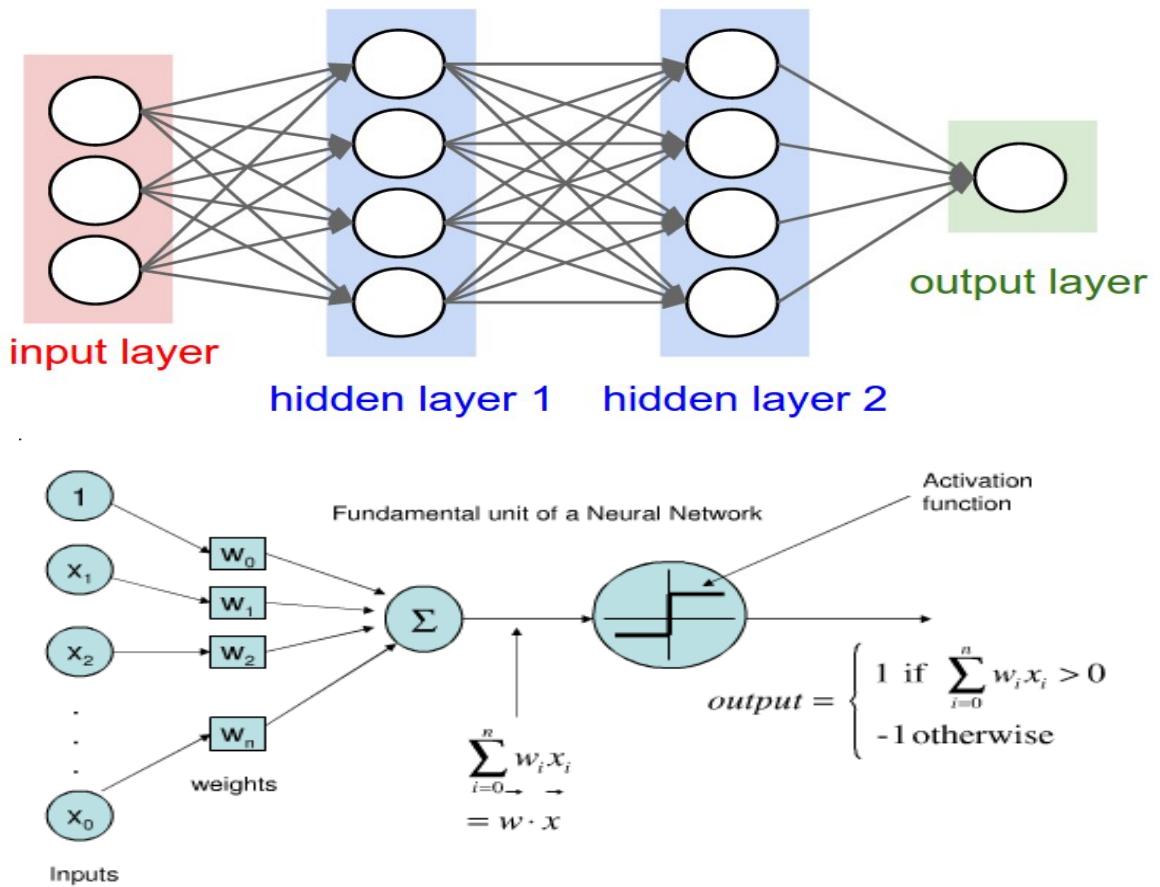
Механико-математический факультет

МГУ им. М. В. Ломоносова

# Современные виды нейросетей

# Нейросетевые вычисления

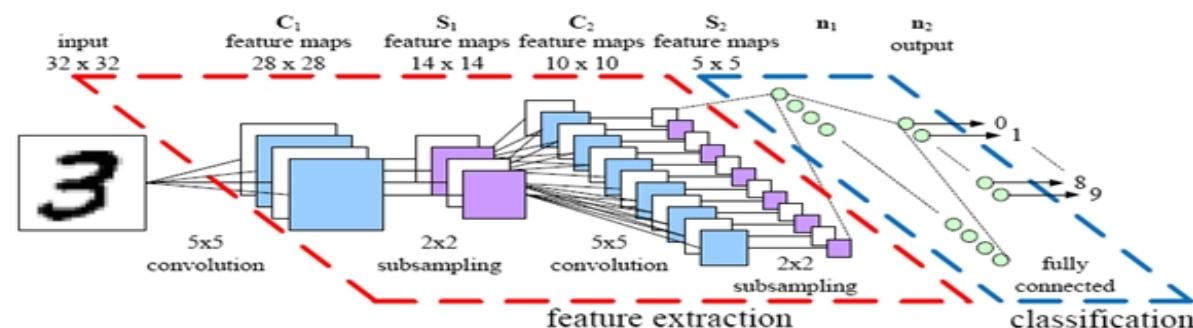
- Главный сегодняшний инструмент для когнитивных технологий
  - Глубокие нейронные сети (Deep Neural Nets, DNN)
- “Глубокие”
  - Множество последовательных вычислительных слоев (десятки или сотни)
- “Нейронные сети”
  - Вычислительная единица NN: ВХОД → ВЫХОД отдаленно напоминает структуру нейрона человеческого мозга



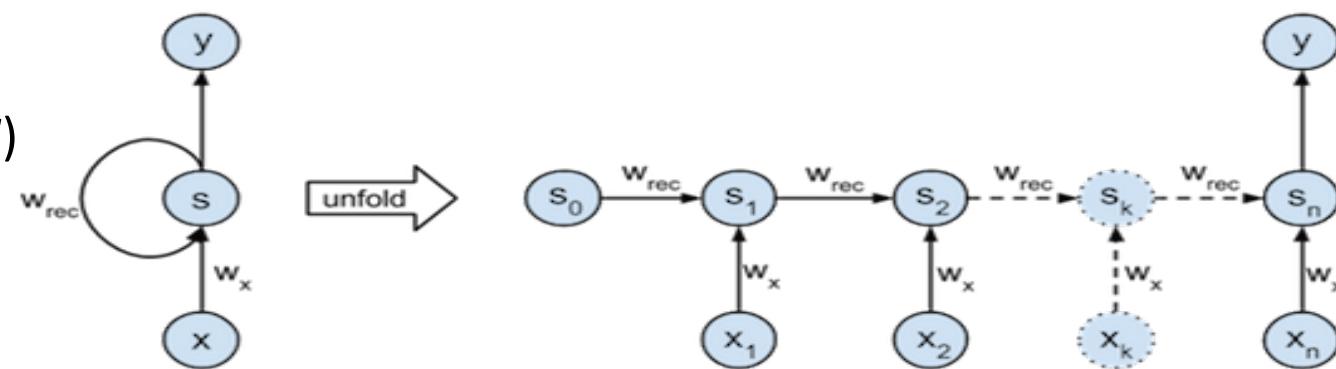
# 2 главных типа нейросетей

- Наиболее важные типы глубоких NN на данный момент

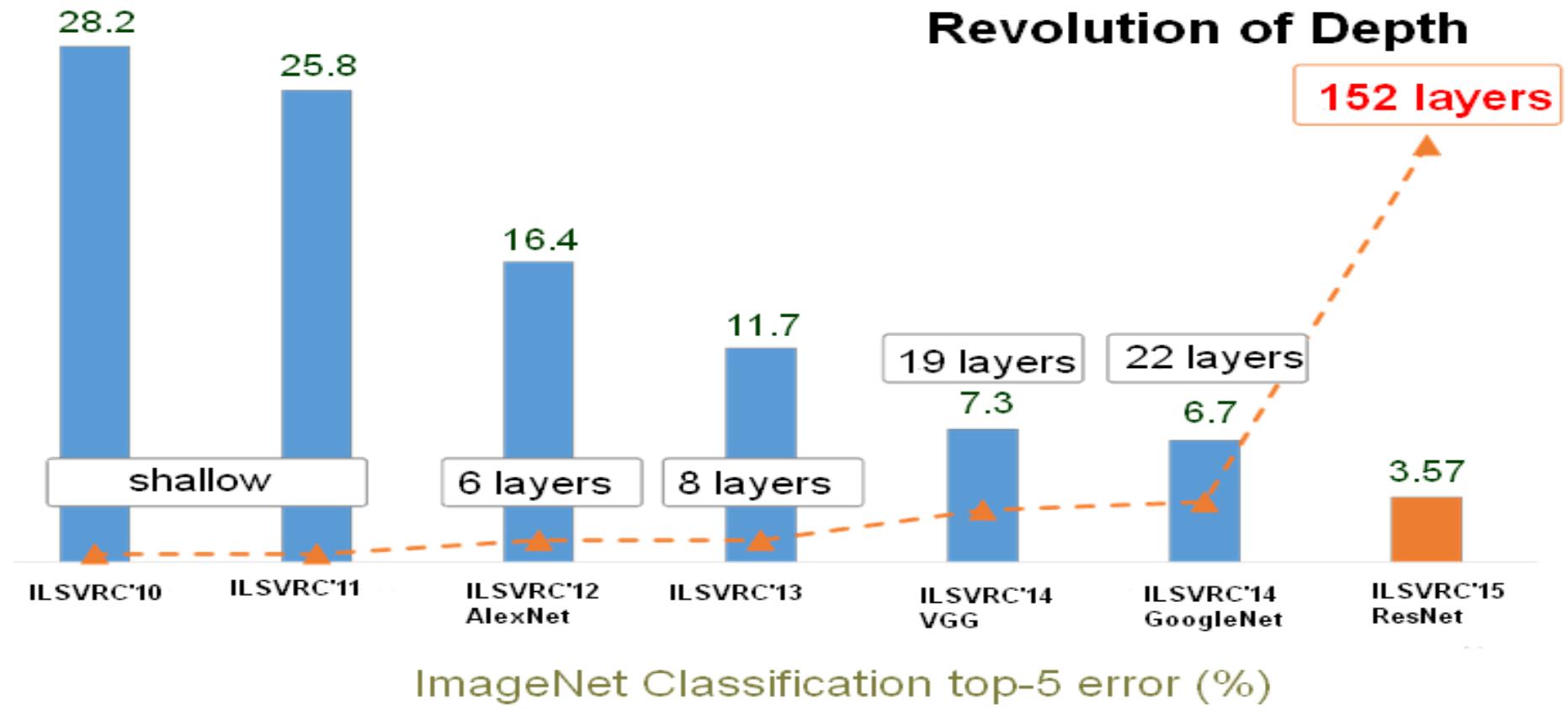
- Сверточная нейросеть прямого распространения (Convolutional NN, CNN)



- Рекуррентная NN (Recurrent NN, RNN)



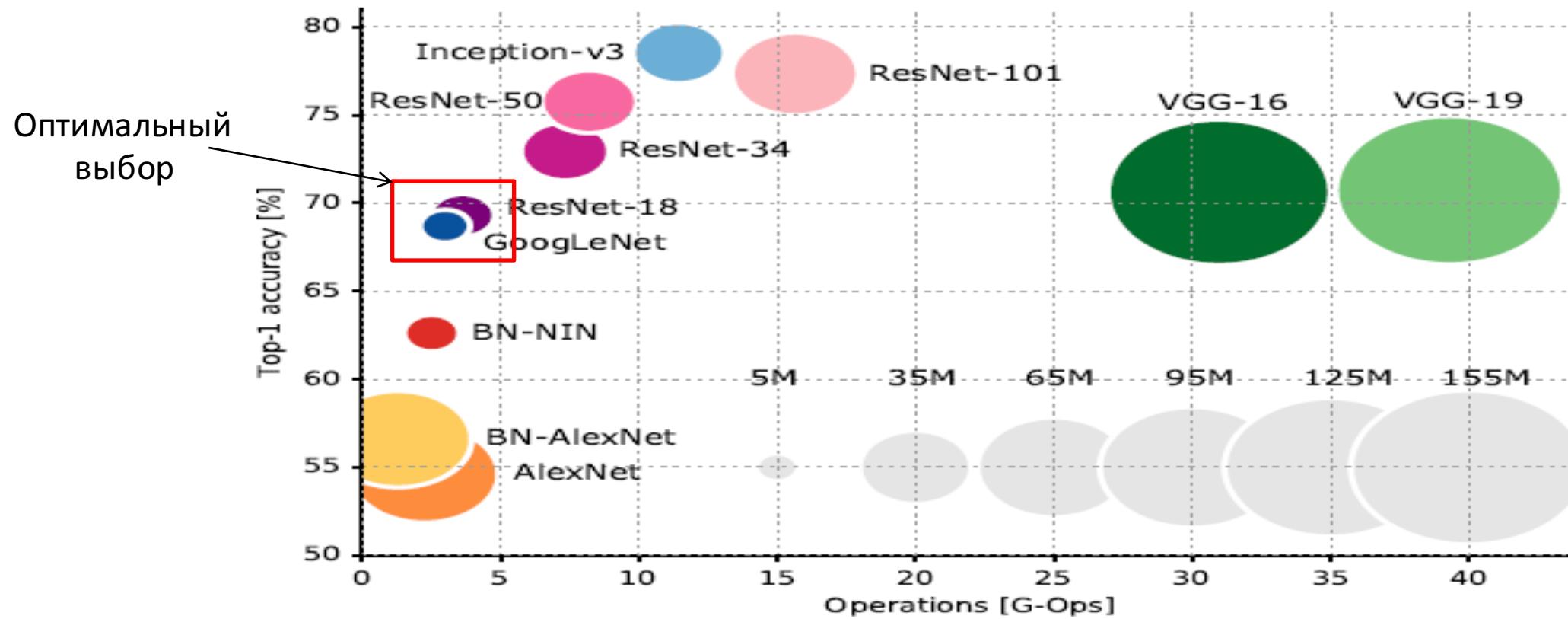
# Эволюция глубоких архитектур



\*ILSVRC = ImageNet<sup>1</sup> Large Scale Visual Recognition Competition:

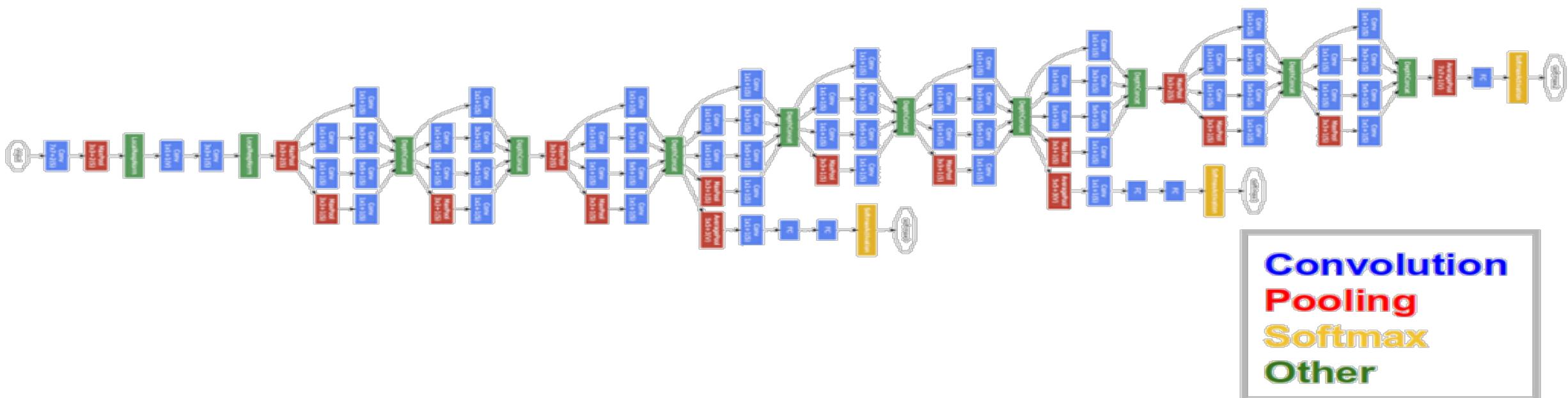
- 50.000 изображений для валидации
- 1.200.000 изображений для обучения

# Зависимость точности от сложности модели NN



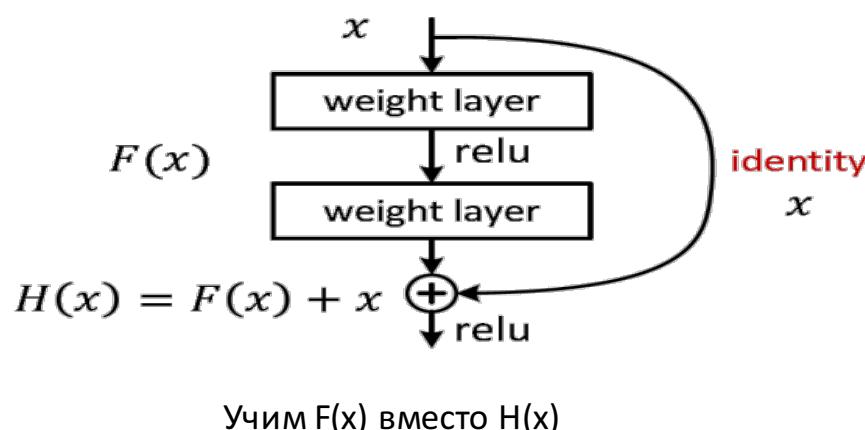
# Пример сверточной нейросети

- Одна из наиболее практических CNN для классификации изображений - это *Inception V1* (также известна как *GoogleNet*<sup>1</sup>): небольшая (десятки мегабайт), но очень точная

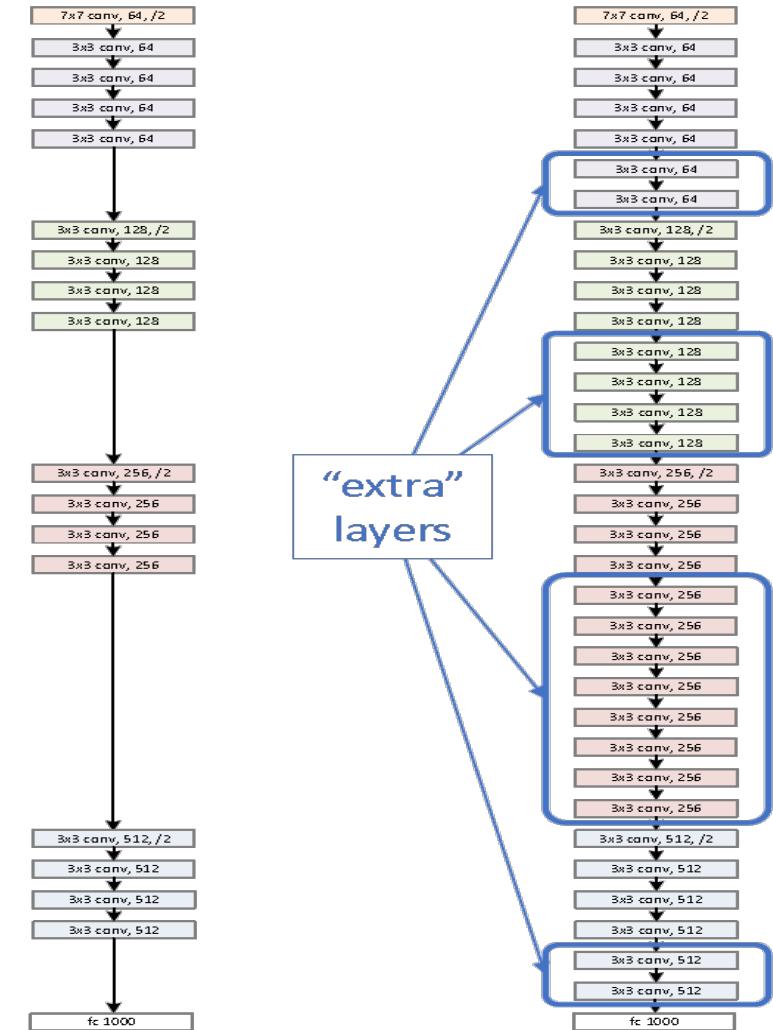


# Еще один пример сверточной нейросети

- Архитектура ResNet<sup>1</sup>
  - Сделана в Microsoft Research
  - Победитель в 2015 г. конкурса классификации картинок
  - CNN с необычной структурой
  - Побила CNN от Google ☺

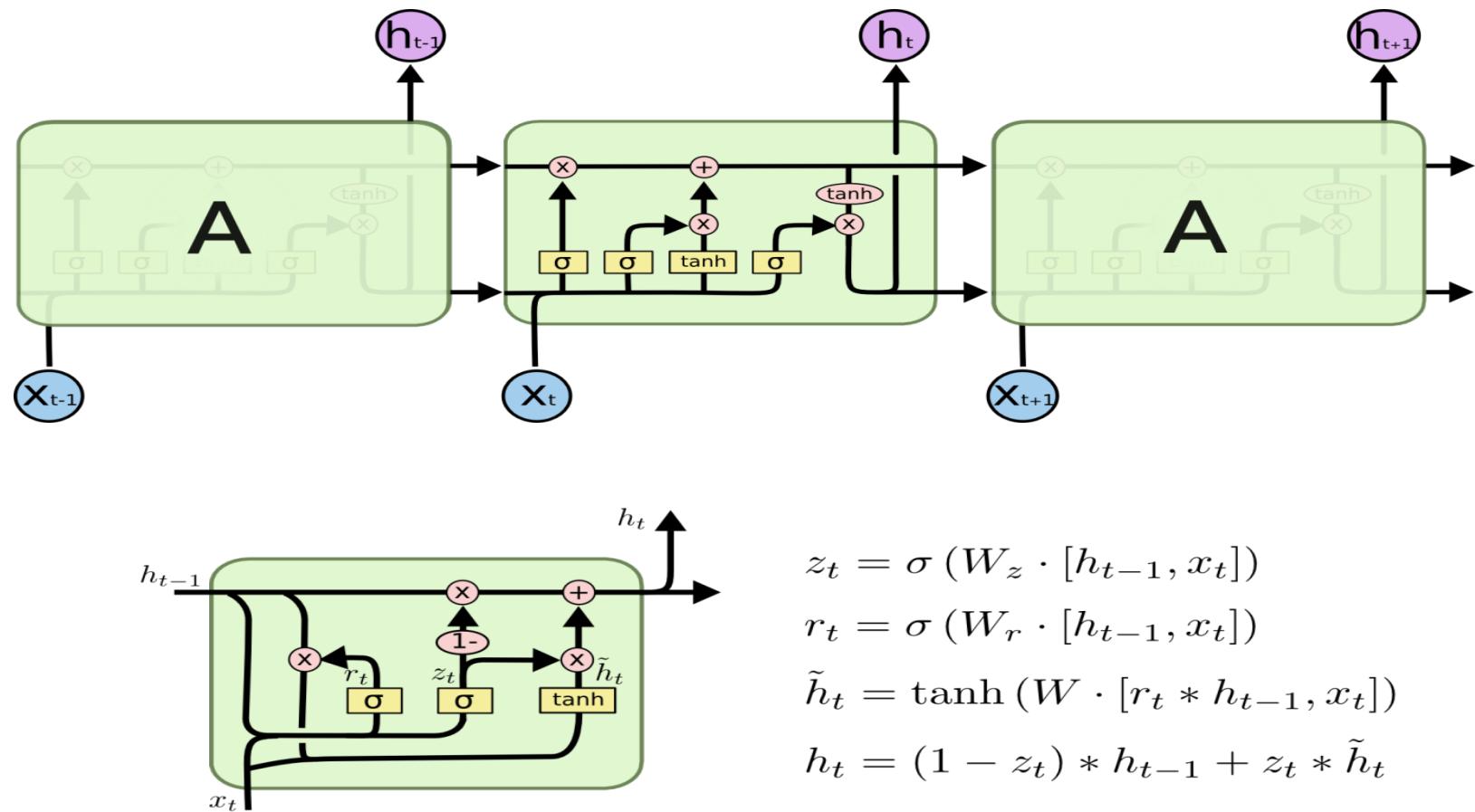


Учим  $F(x)$  вместо  $H(x)$



# Пример рекуррентной нейросети

- Наиболее полезные RNN для обработки текста/речи
  - Долгая краткосрочная память (Long short-term memory, *LSTM*<sup>1</sup>)
  - И ее улучшенная версия – запираемая рекуррентная ячейка (Gated recurrent unit, *GRU*<sup>2</sup>)



[1] S. Hochreiter, J. Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780

[2] J. Chung et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." 2014, <https://arxiv.org/abs/1412.3555>

# Примеры использования нейросетей

# Реальные продукты, использующие глубокие нейросети

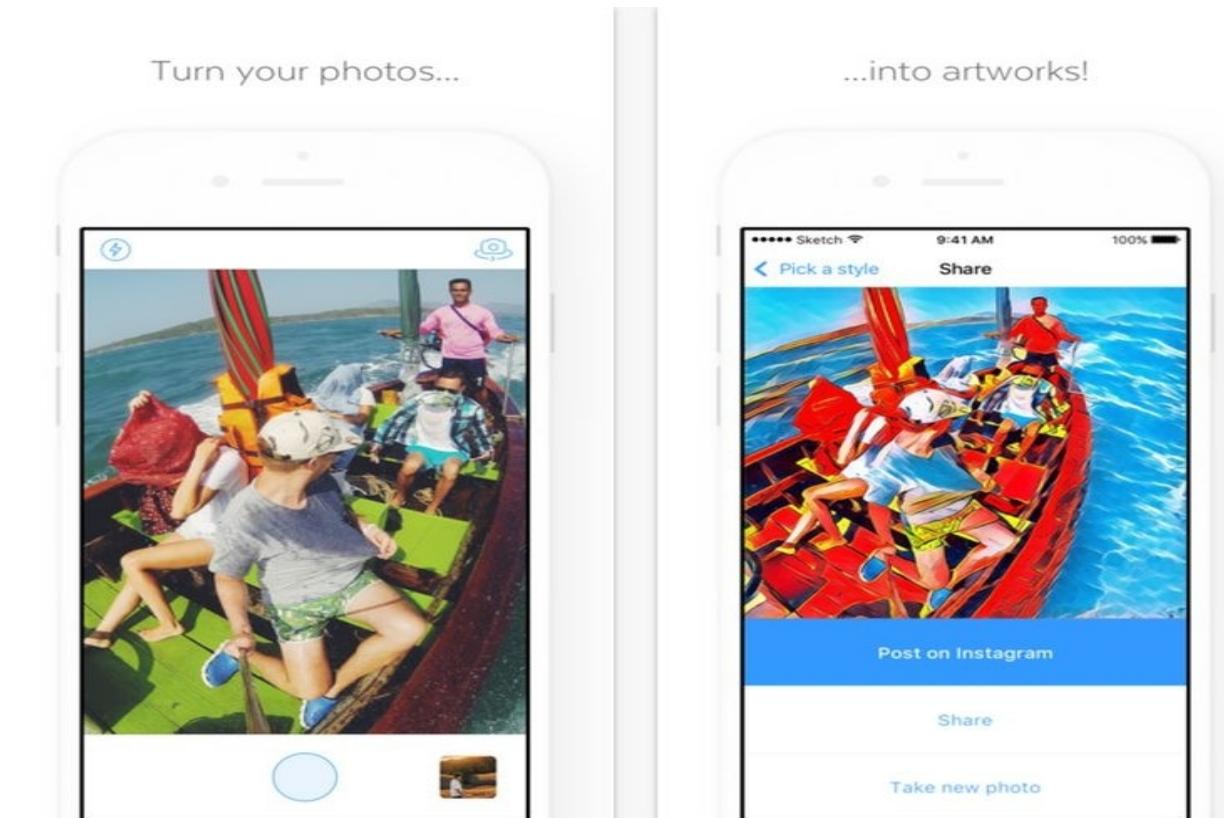
- DeepFace<sup>1</sup>
  - Разработан группой Яна Лекуна (Yann LeCun) в Facebook AI lab
    - Ян Лекун – отец-основатель современных CNN
  - Автоматически распознает людей по их лицам на фотографиях



# Реальные продукты, использующие глубокие нейросети

- Стилизация “Призма”<sup>1</sup>

- Технология разработана бывшим сотрудником Mail.ru Group
- Реализация статьи 2015 г.<sup>2</sup>
- Используется CNN со специальной функцией потерь (матрица Грамма)



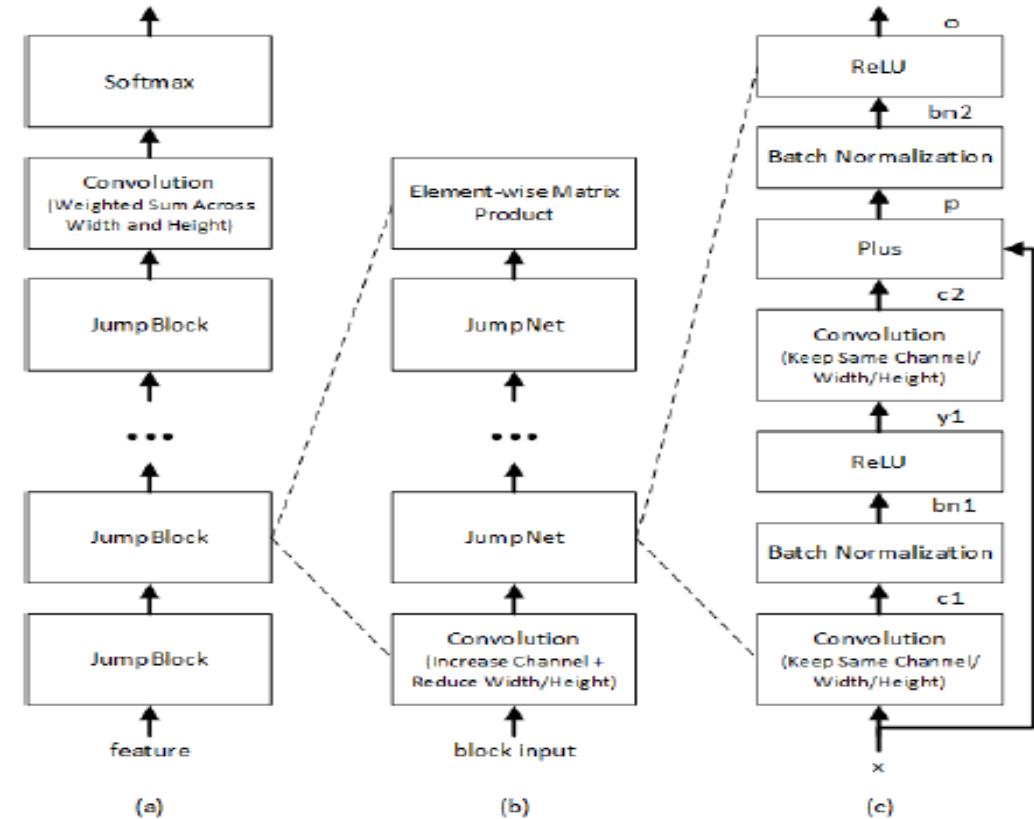
[1] <http://prisma-ai.com/>

[2] L. Gatys et al. "A Neural Algorithm of Artistic Style." 2015, <https://arxiv.org/abs/1508.06576>

# Реальные продукты, использующие глубокие нейросети

- Система LACE для автоматического распознавания речи<sup>1</sup>
  - Сделана в Microsoft Research
  - Нейросеть с временной задержкой (Time delay neural network, TDNN)
  - Иногда справляется лучше человека

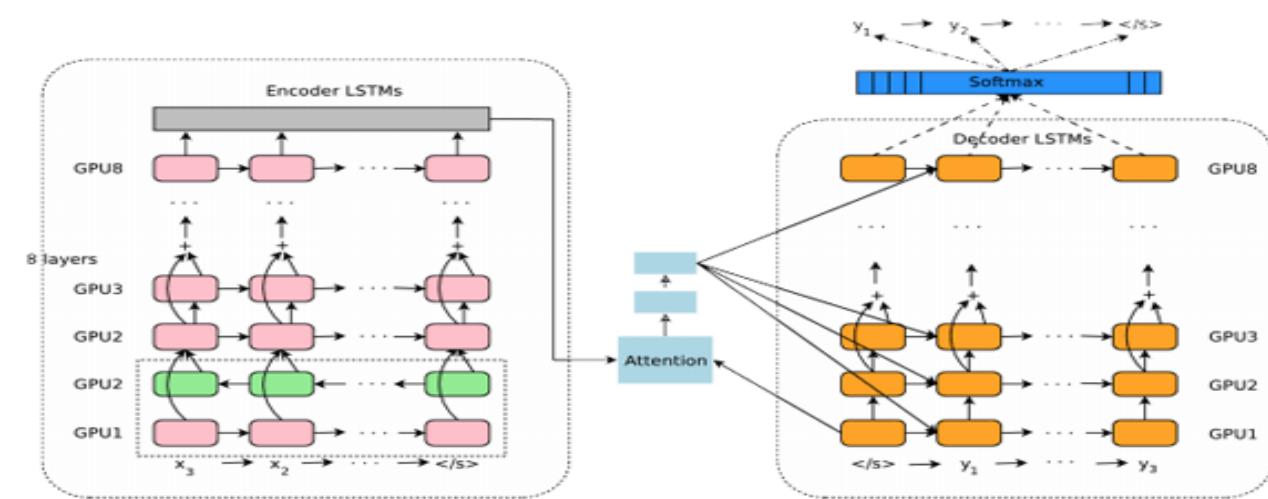
Model	N-gram LM		Neural net LM	
	CH	SWB	CH	SWB
Povey et al. [54] LSTM	15.3	8.5	-	-
Saon et al. [51] LSTM	15.1	9.0	-	-
Saon et al. [51] system	13.7	7.6	12.2	6.6
2016 Microsoft system	13.3	7.4	11.0	5.8
Human transcription			11.3	5.9



# Реальные продукты, использующие глубокие нейросети

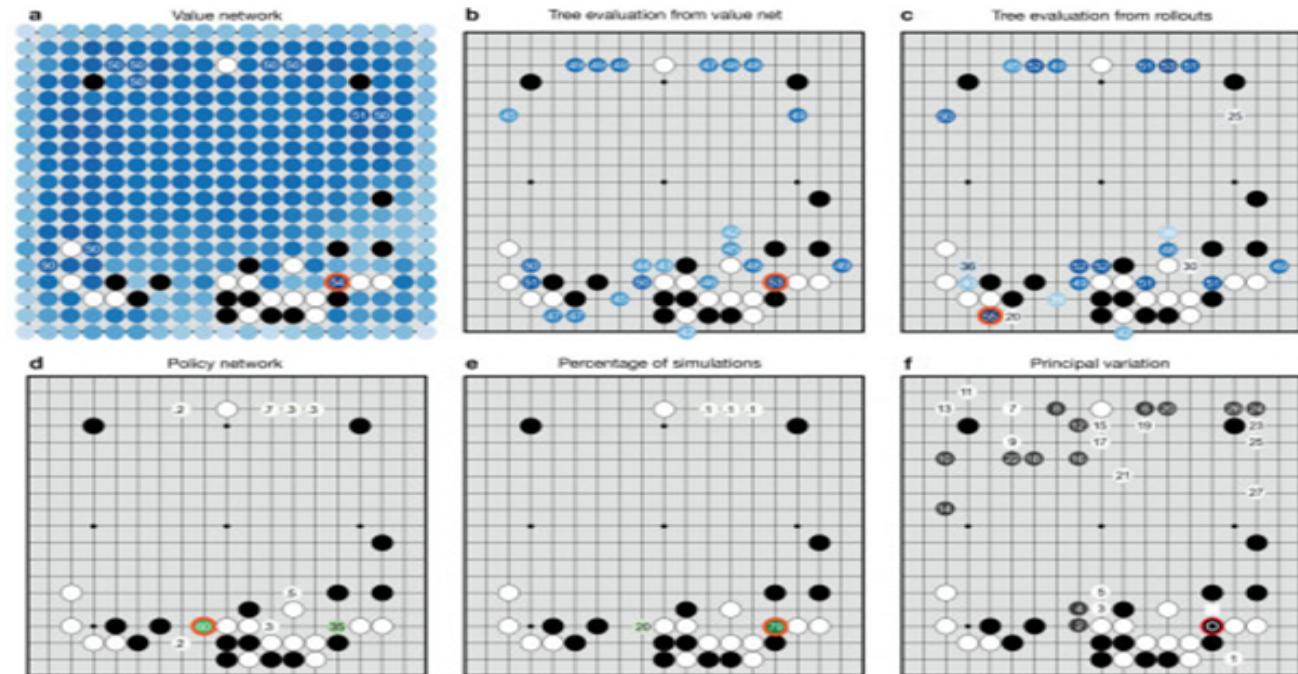
- Система автоматического перевода GNMT<sup>1</sup>
  - Google's Neural Machine Translation
  - 8 двунаправленных LSTM
  - Для некоторых пар языков (например, english/spanish) работает на уровне переводчиков ☺

	PBMT	GNMT	Human
English → Spanish	4.885	5.428	5.504
English → French	4.932	5.295	5.496
English → Chinese	4.035	4.594	4.987
Spanish → English	4.872	5.187	5.372
French → English	5.046	5.343	5.404
Chinese → English	3.694	4.263	4.636



# Реальные продукты, использующие глубокие нейросети

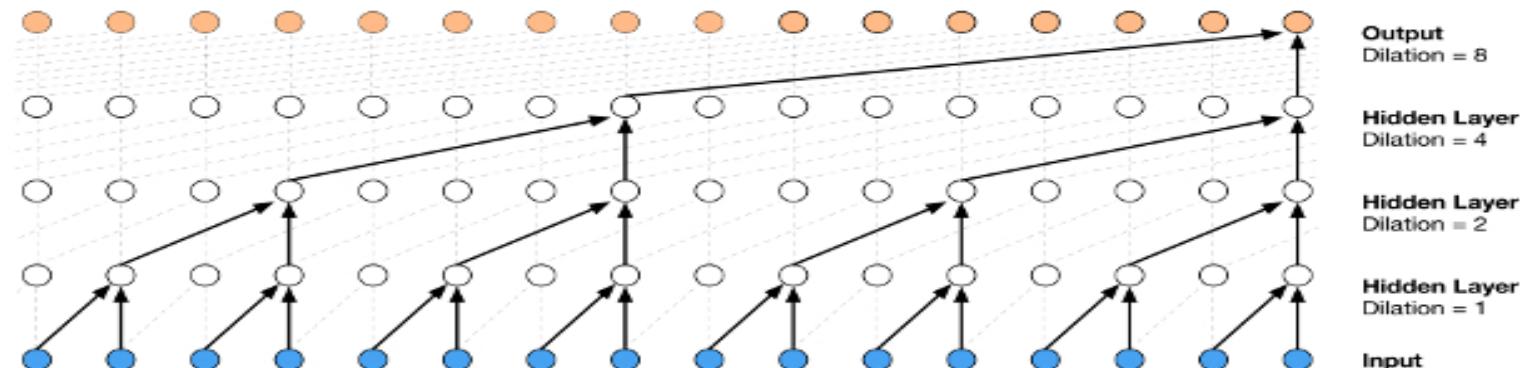
- DeepMind
- 2015: AlphaGo
  - Последняя игра с полной информацией, в которой человек был побежден компьютером



# Реальные продукты, использующие глубокие нейросети

- DeepMind
- 2016: WaveNet<sup>1</sup>
  - Система речевого синтеза
  - Лучшая система на данный момент

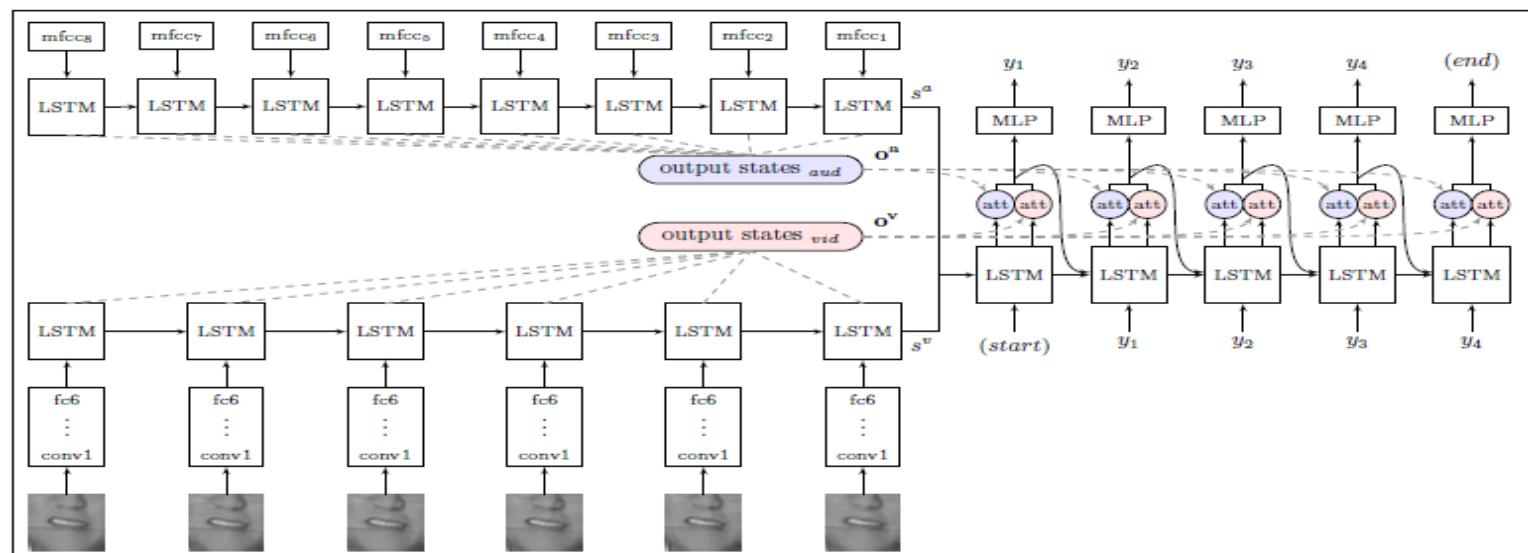
Speech samples	Subjective 5-scale MOS in naturalness	
	North American English	Mandarin Chinese
LSTM-RNN parametric	$3.67 \pm 0.098$	$3.79 \pm 0.084$
HMM-driven concatenative	$3.86 \pm 0.137$	$3.47 \pm 0.108$
<b>WaveNet (L+F)</b>	<b><math>4.21 \pm 0.081</math></b>	<b><math>4.08 \pm 0.085</math></b>
Natural (8-bit $\mu$ -law)	$4.46 \pm 0.067$	$4.25 \pm 0.082$
Natural (16-bit linear PCM)	$4.55 \pm 0.075$	$4.21 \pm 0.071$



# Реальные продукты, использующие глубокие нейросети

- DeepMind<sup>1</sup> + Оксфорд<sup>2</sup>
- 2016: WLAS (Watch, Listen, Attend and Spell)
  - Система чтения по губам
  - Лучше человека 😊

Method	SNR	CER	WER	BLEU <sup>†</sup>
<b>Lips only</b>				
Professional <sup>‡</sup>	-	58.7%	73.8%	23.8
WAS	-	59.9%	76.5%	35.6
WAS+CL	-	47.1%	61.1%	46.9
WAS+CL+SS	-	42.4%	58.1%	50.0
WAS+CL+SS+BS	-	39.5%	50.2%	54.9

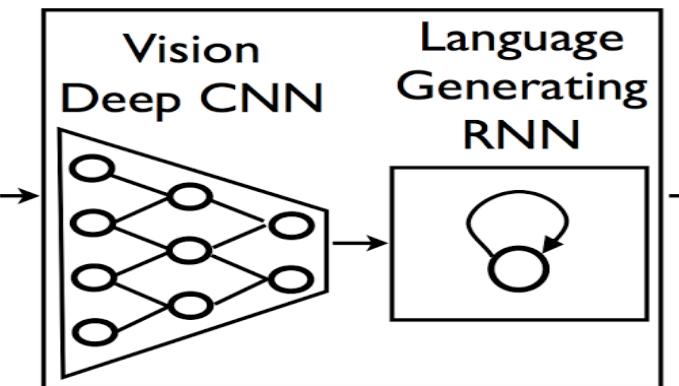


[1] J. S. Chung et al. "Lip Reading Sentences in the Wild." 2016, <https://arxiv.org/abs/1611.05358>

[2] Y. M. Assael et al. "LipNet: Sentence-level Lipreading." 2016, <https://arxiv.org/abs/1611.01599>

# Исследования в области глубоких нейросетей

- Генерация подписей к изображениям by Andrej Karpathy



**A group of people shopping at an outdoor market.**

**There are many vegetables at the fruit stand.**

# Исследования в области глубоких нейросетей

- Раскрашивание черно-белых изображений<sup>2</sup>
- CNN

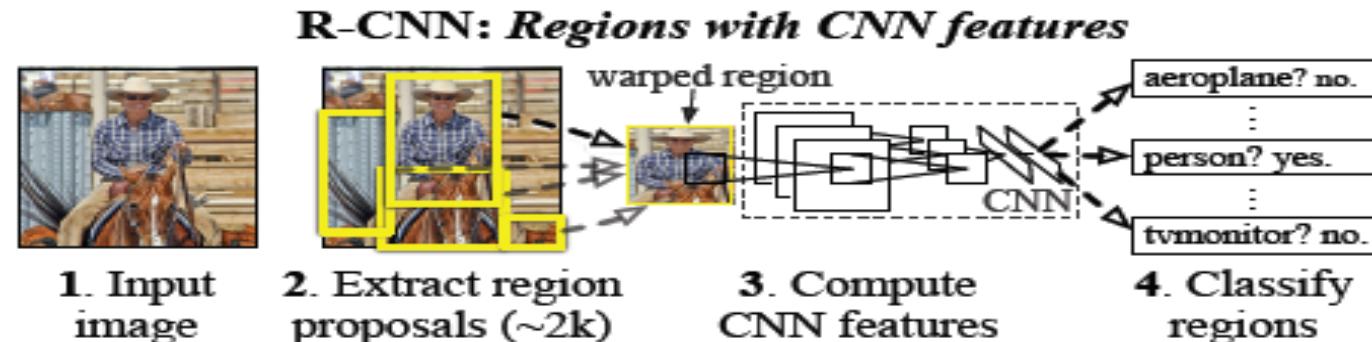


[1] <http://www.berkeley.edu/>

[2] R. Zhang et al. "Colorful Image Colorization." 2016, <https://arxiv.org/abs/1603.08511>

# Исследования в области глубоких нейросетей

- Университет Беркли
  - Классический подход для детекции объектов R-CNN<sup>1,2</sup>
  - Разработан by Ross Girshick (перешел в Facebook AI Research)
  - В момент изобретения выиграл все НЕ-нейросетевые подходы в соревновании по детекции VOC (и на данный момент его идеи используются во всех системах детекции)

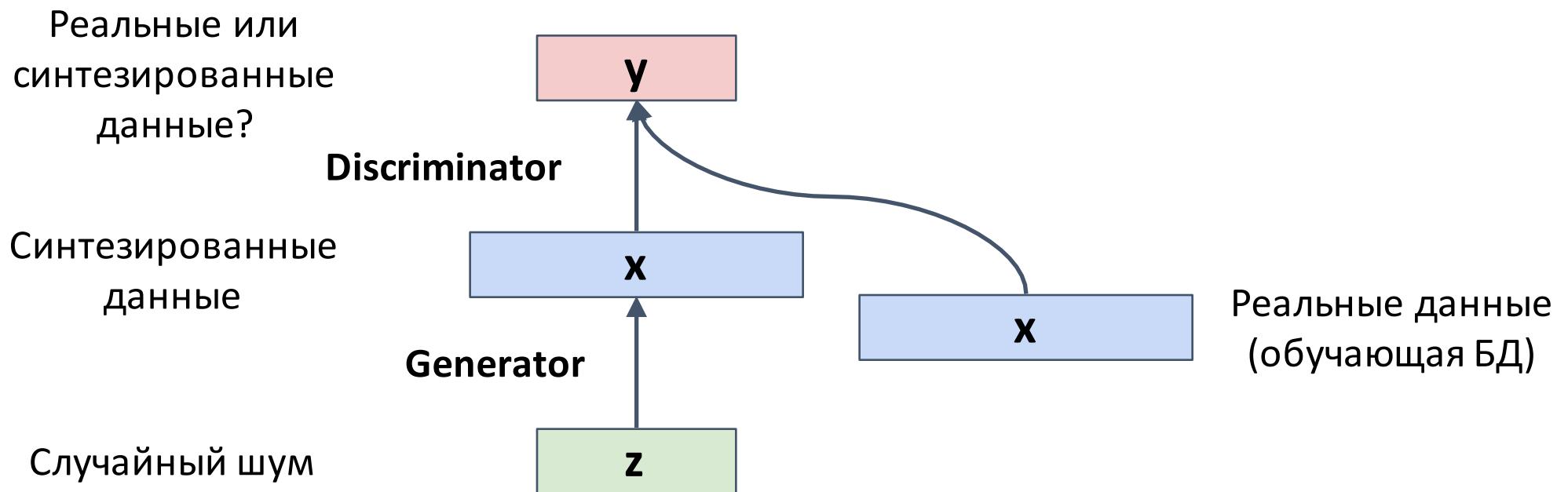


[1] R. Girshick. "Fast R-CNN." 2015, <https://arxiv.org/abs/1504.08083>

[2] <https://github.com/rbgirshick/py-faster-rcnn>

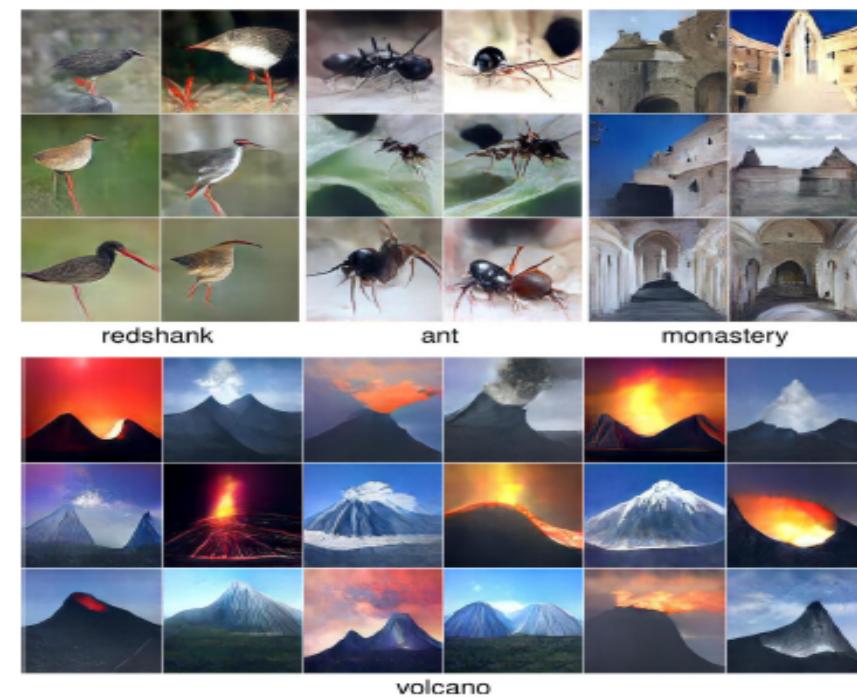
# GAN

- Генеративные состязательные сети (Generative Adversarial Networks, GAN)
  - Обучение Generator и Discriminator совместно, итерация за итерацией



# Пример применения GAN

- PPGN – генеративная модель с размером картинки (227x227)



# Пример применения GAN

- Сверх-разрешение с помощью GAN

original



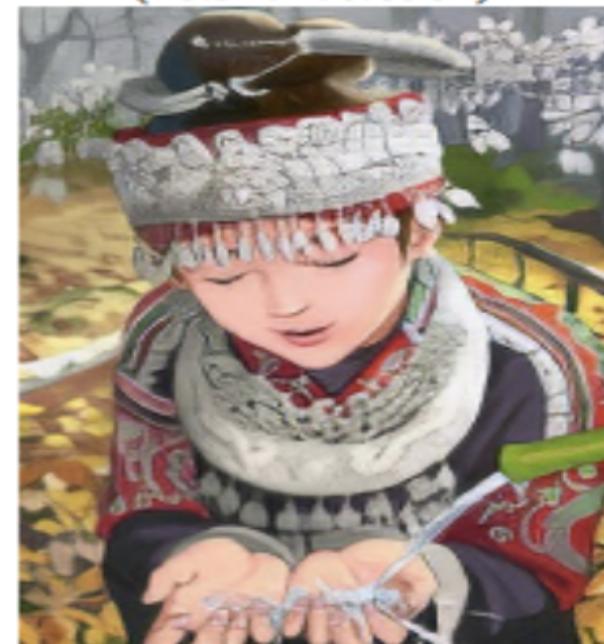
bicubic  
(21.59dB/0.6423)



SRResNet  
(23.44dB/0.7777)



SRGAN  
(20.34dB/0.6562)



# Пример применения GAN

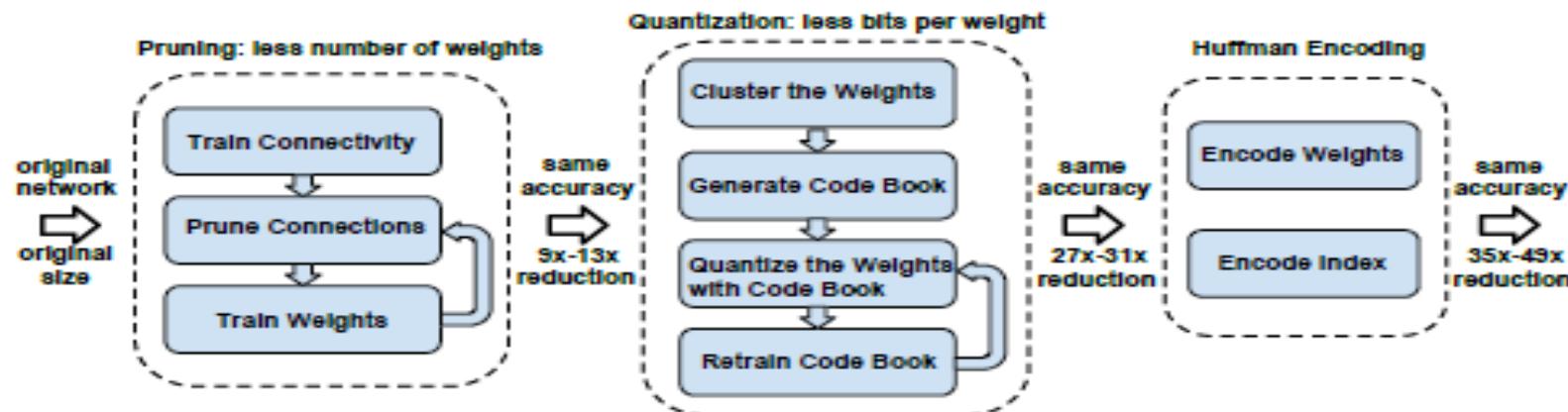
- Качество практически неотличимо<sup>1</sup> от реальных фотографий
- Лица: 1024x1024



[1] Karras, Tero, et al. "Progressive growing of gans for improved quality, stability, and variation." arXiv preprint arXiv:1710.10196 (2017).

# Нейросети: ближе к практике

- Статья Song Han о «глубоком сжатии»<sup>1</sup>
- Сжатие CNN с помощью квантования и прореживания
- Сжатие в 40 раз без потери качества на классической CNN типа AlexNet (на топовых сегодняшних сетях сжатие похуже)

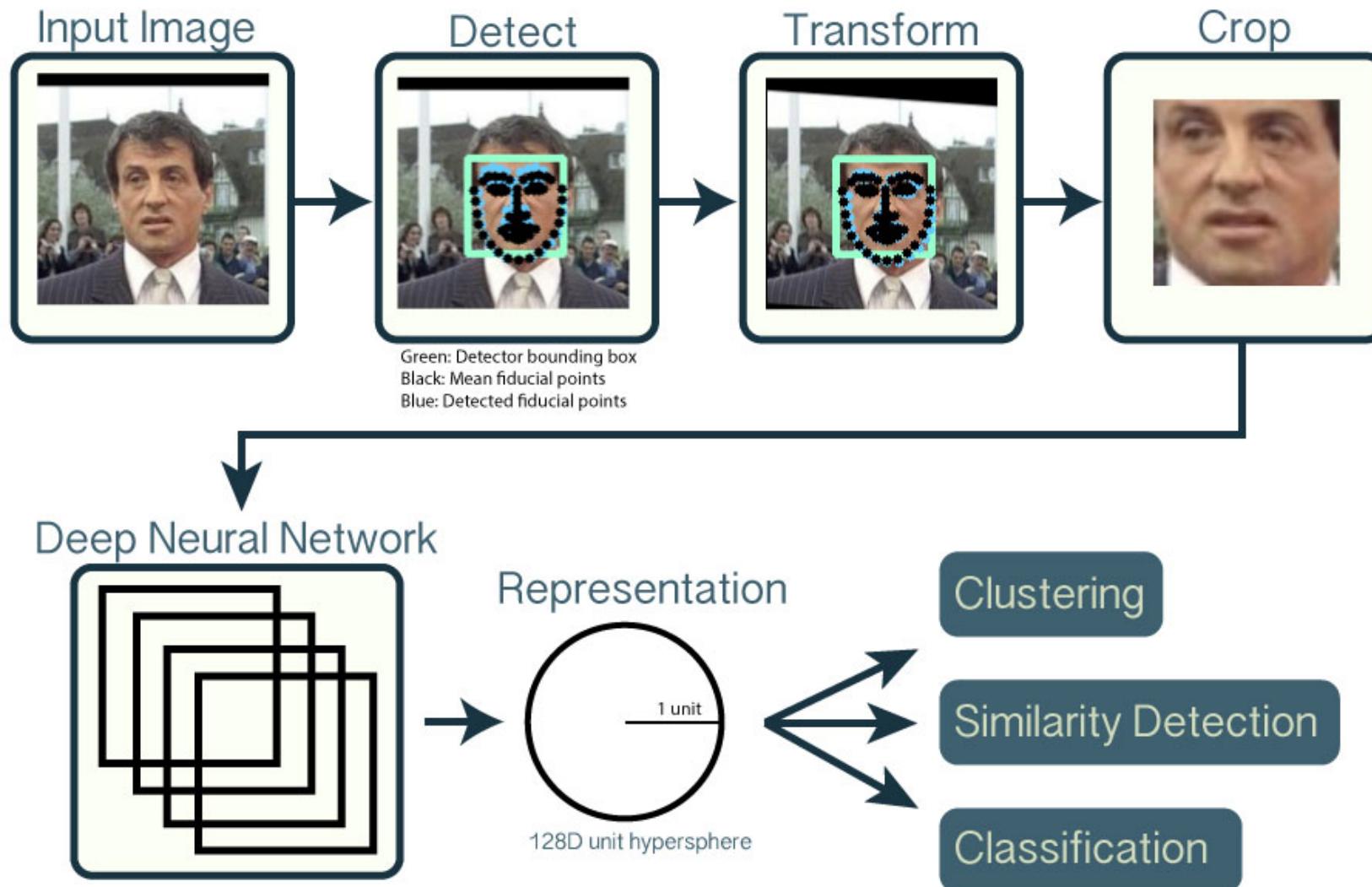


# Фреймворки для глубоко обучения

- Caffe - <http://caffe.berkeleyvision.org/>
  - Наиболее предпочтителен в академической среде
  - Разрабатывается в Университете Беркли: UC Berkeley Vision and Learning Center
- TensorFlow - <https://www.tensorflow.org/>
  - Самый быстроразвивающийся фреймворк для глубокого обучения
  - Разрабатывается в Google Brain
- CNTK – <https://www.microsoft.com/en-us/research/product/cognitive-toolkit/>
  - Cognitive Toolkit
  - Разрабатывается в Microsoft
- Torch - <http://torch.ch/>
  - Основан на экзотическом языке программирования Lua
  - Используется в Facebook, IBM и Yandex
- OpenAI и обучение с подкреплением
  - <https://universe.openai.com> - интерактивная платформа-песочница для обучения с подкреплением
  - <https://gym.openai.com> - набор готовых окружений для обучения агентов

# Распознавание лиц

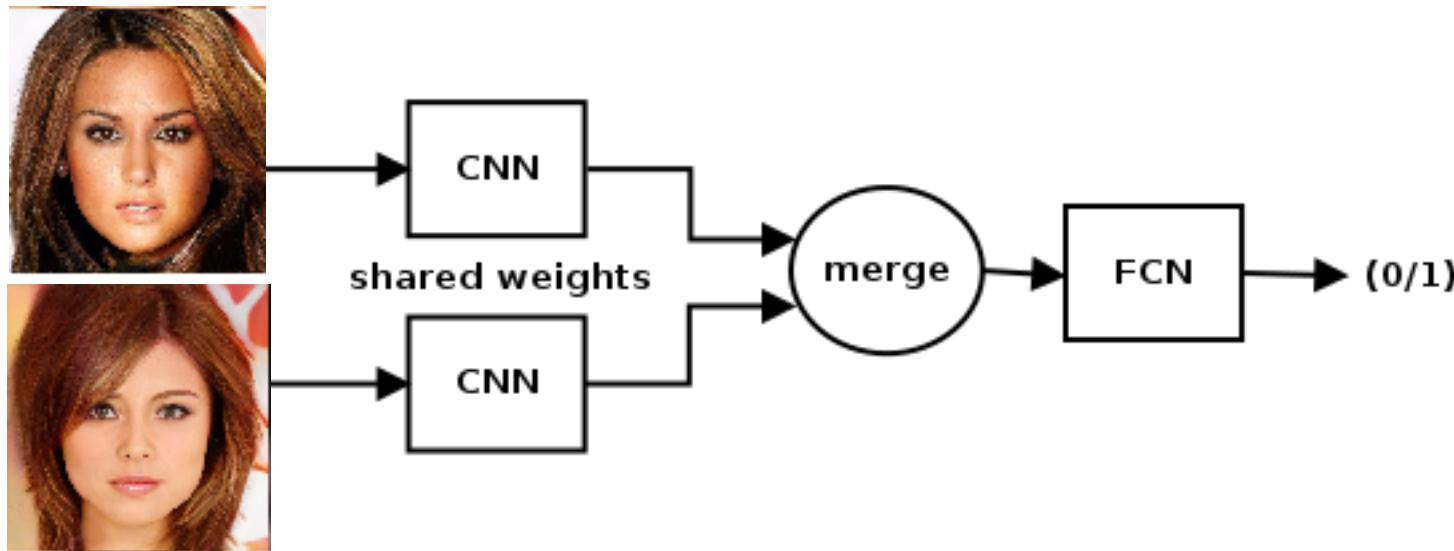
# Современный подход к распознаванию лиц



# Главные задачи в распознавании лиц

- Задачи
  - Верификация
    - Определение того, являются ли поданные на вход картинки изображениями одного и того же человека
  - Идентификация
    - Нахождение в базе данных ID человека, изображение которого подано на вход
  - Проекция (embedding)
    - Нахождение вектора представления лица размерности, намного меньшей входной

# Верификация

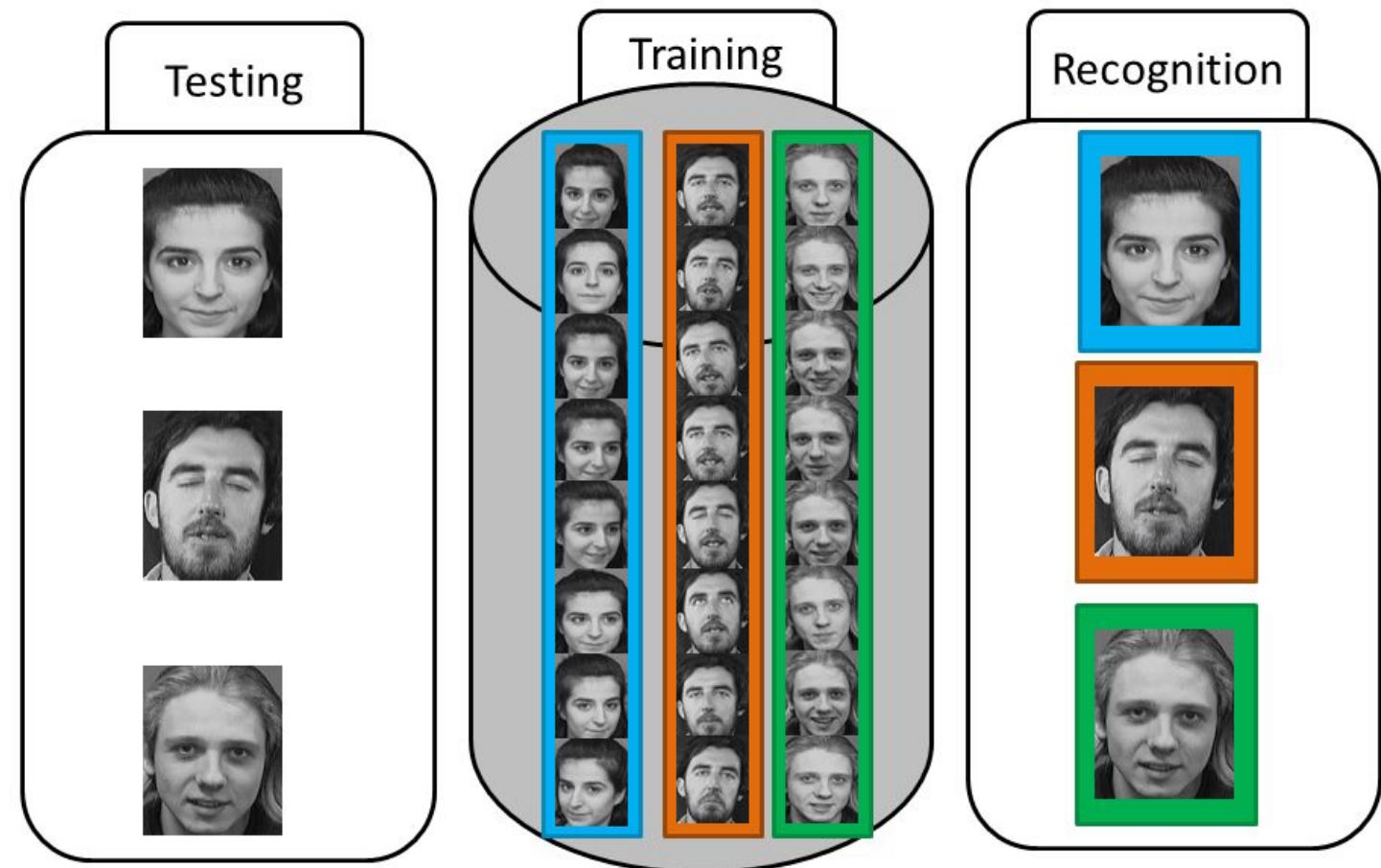


Самый простой метод – использование т.н. Сиамских сетей. Вектор признаков вытаскивает одна и та же сеть, после чего применяется классификатор для итогового решения

# Идентификация

- Если грубо, то это ОЧЕНЬ большой классификатор
  - Количество классов – число разных людей для распознавания

## The Problem



# Проекция

- Так или иначе используется во всех задачах, связанных с лицами
- Обычно элементы вектора представлений редко имеют содержательный смысл по отдельности

