

Сверточные нейронные сети в задачах компьютерного зрения



Кафедра
технологий
проектирования
сложных
технических
систем

Основные направления в CV

Классификация

Решается задача
принадлежности
изображения одному из
классов



Основные направления в CV

Сегментация

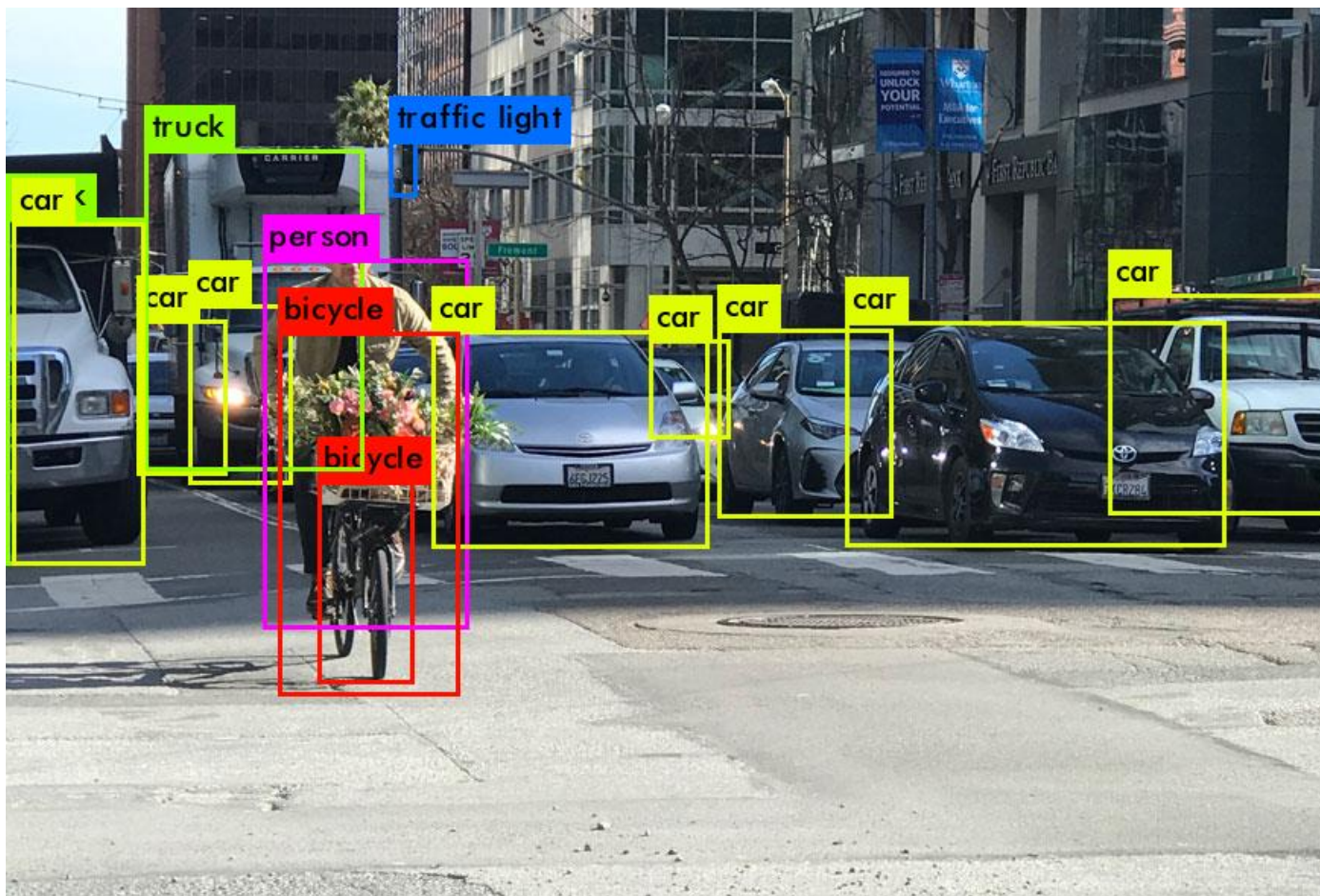
Решается задача классификации, но попиксельной



Основные направления в CV

Детекция

Решается задача
распознавания - найти и
классифицировать



Основные направления в CV

Трекинг и видео-аналитика

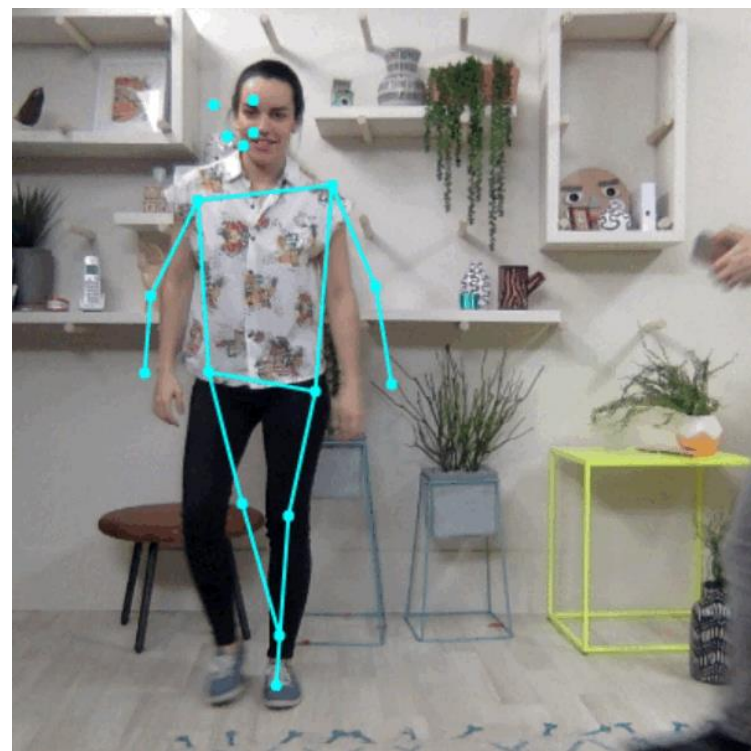
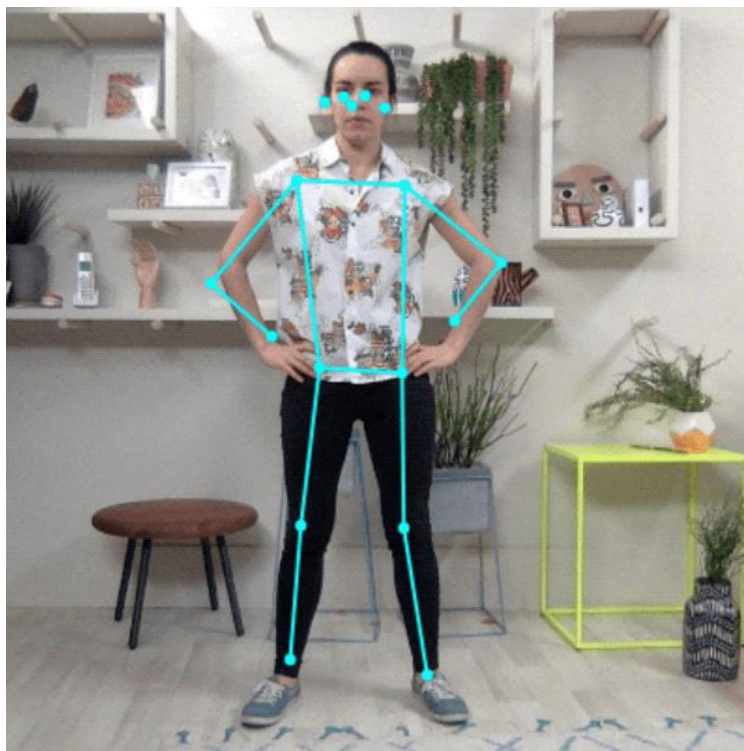
Распознавание объектов в
видеопотоке



Основные направления в CV

Оценка позы и взгляда человека

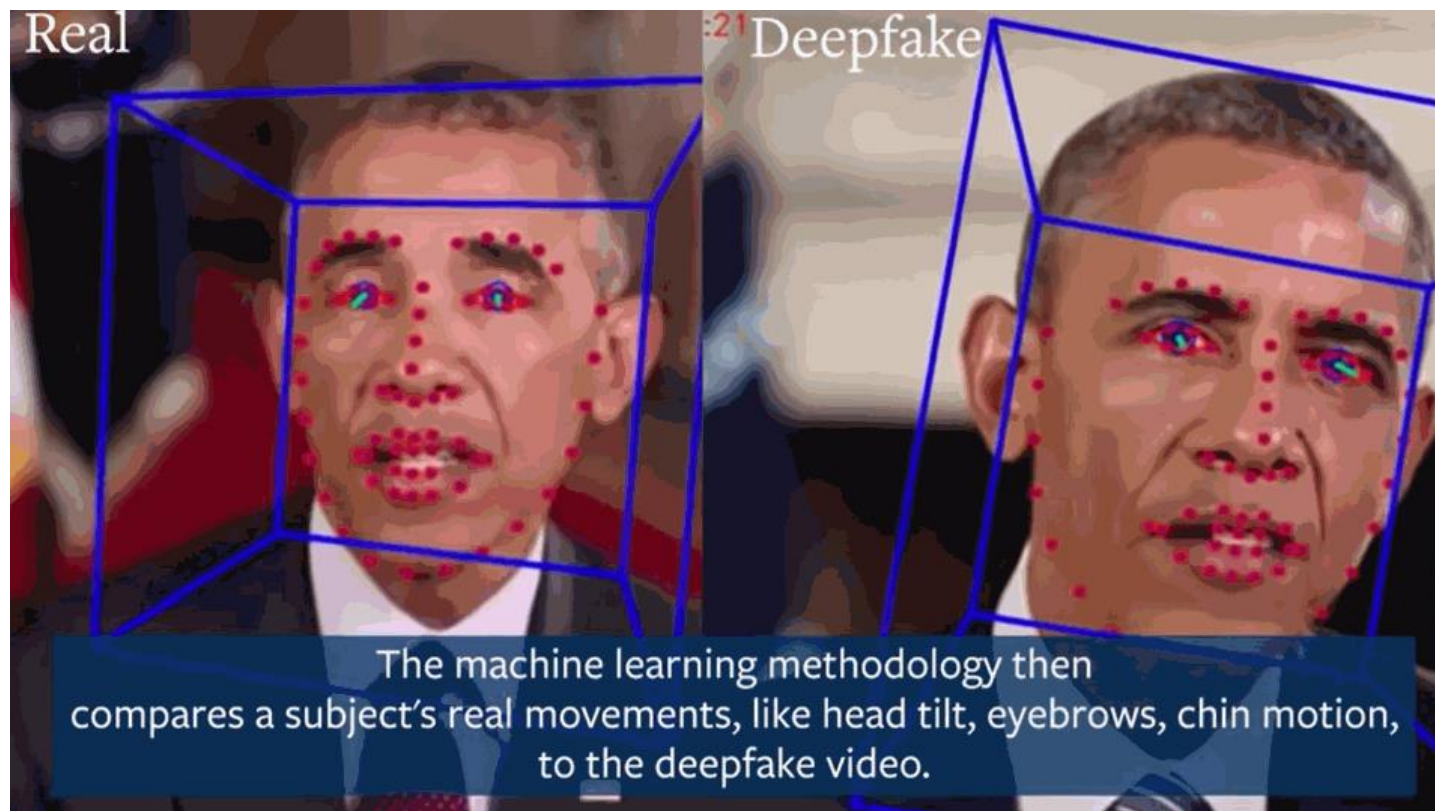
Решается задача распознавания ключевых точек, которые описывают позы человека, его положение, ориентацию в пространстве



Основные направления в CV

Биометрия

Решается задача классификации на изображении или видеопотоке: фейк или нет



Основные направления в CV

Построение карты глубины

Решается задача построения карты расстояний до наблюдаемых объектов



Input video



Our depth predictions

Основные направления в CV

Улучшение и восстановление изображений

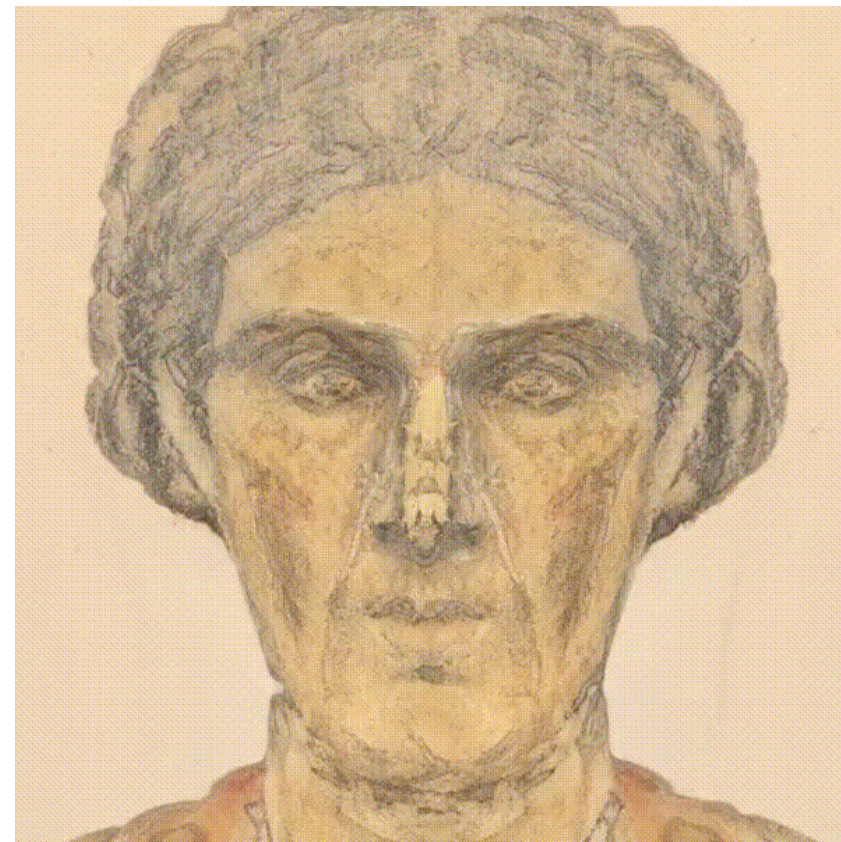
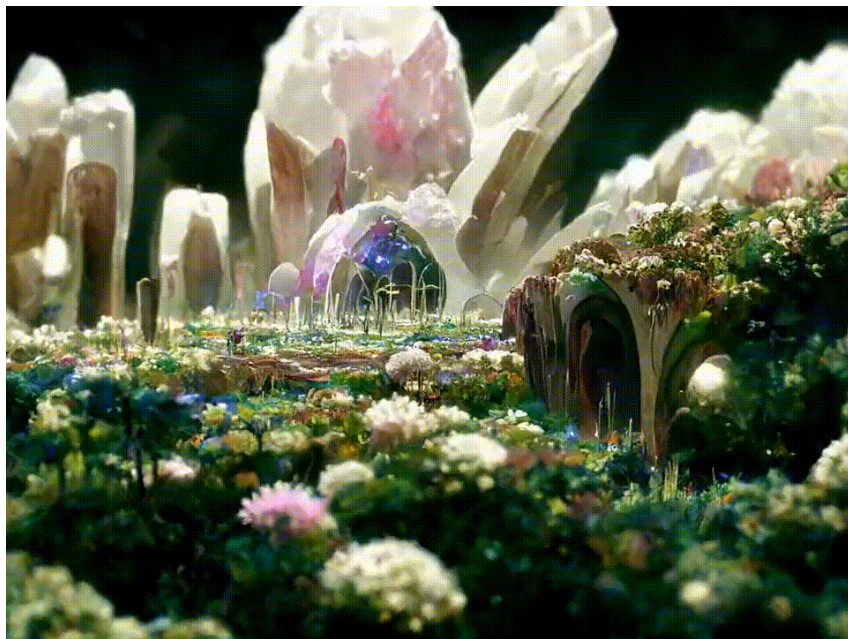
Поиск дефектов и
коррекция изображения,
преобразование стилей



Основные направления в CV

Генерация изображений

Синтез новых
изображений



Основные направления в CV

Рендеринг

VR и нейро рендеринг



Основные направления в CV

Построение 3D Объектов

Решается задача построения объемного объекта (obj-файл) по двумерному изображению



Задача верификации

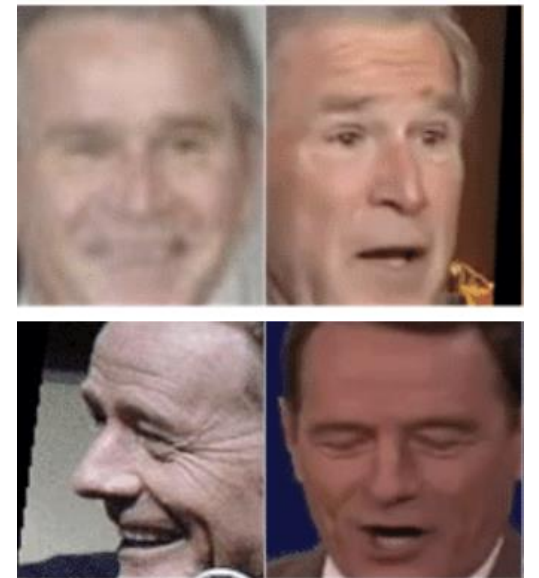
Постановка задачи – на заданных двух изображениях один и тот же объект?

- Алгоритм должен иметь класс «unseen»
- Нерепрезентативная выборка (мало данных, дисбаланс классов)

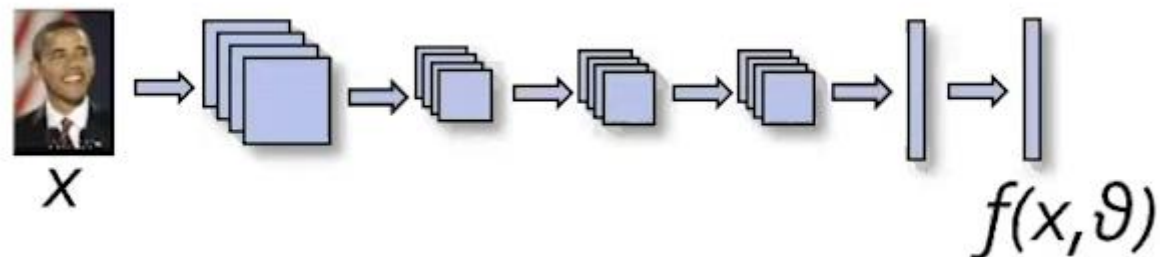
Face recognition dataset (MSRA-CF)

Re-identification dataset (ViPER)

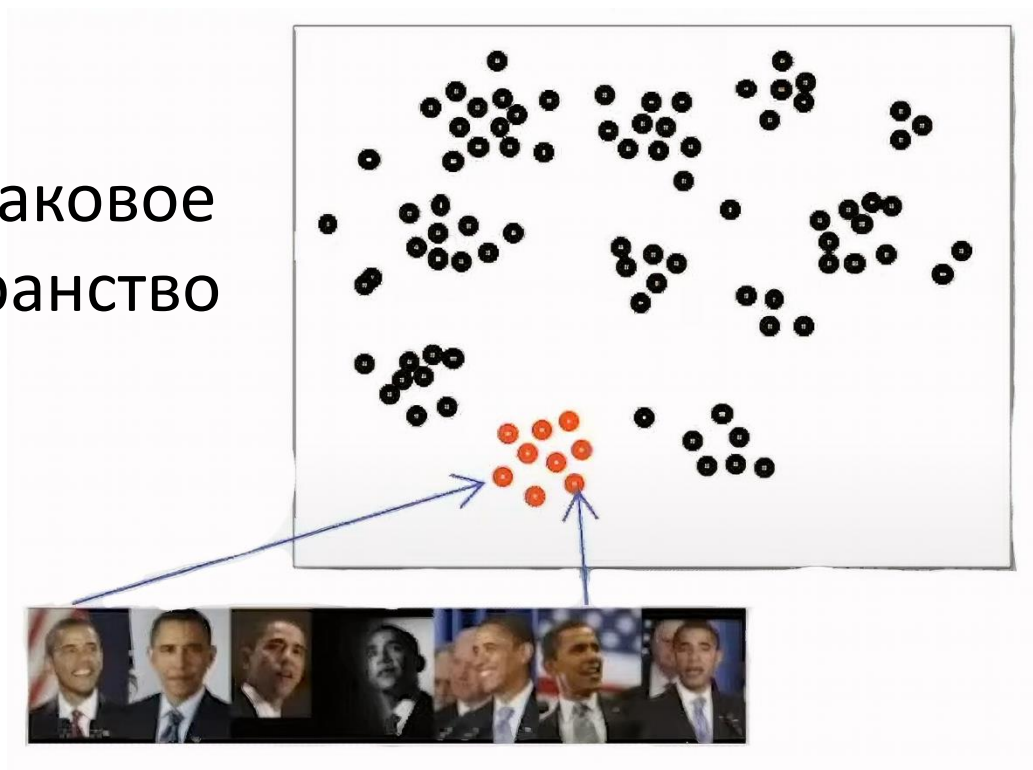
Open world dataset



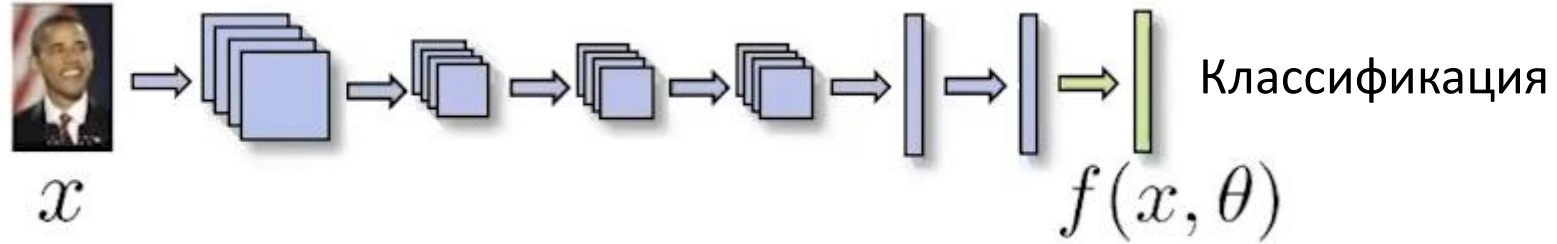
Задача верификации как embedding learning



Признаковое
пространство

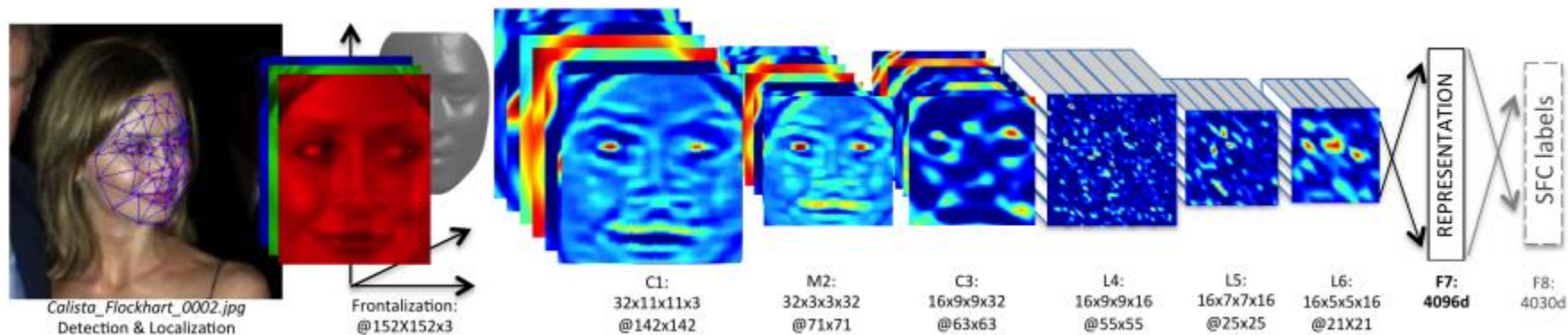


Задача верификации как embedding learning

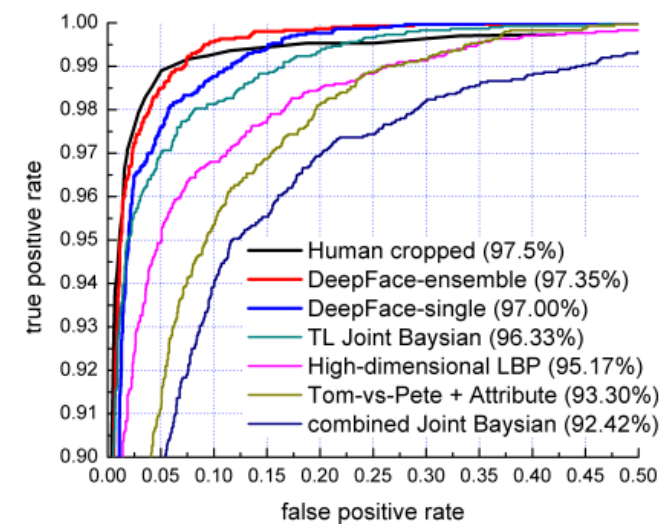


- Обучение на задаче классификации
- Чем больше данных, тем лучше результат
- Классы во время обучения могут быть в виде прототипов классов на тестировании

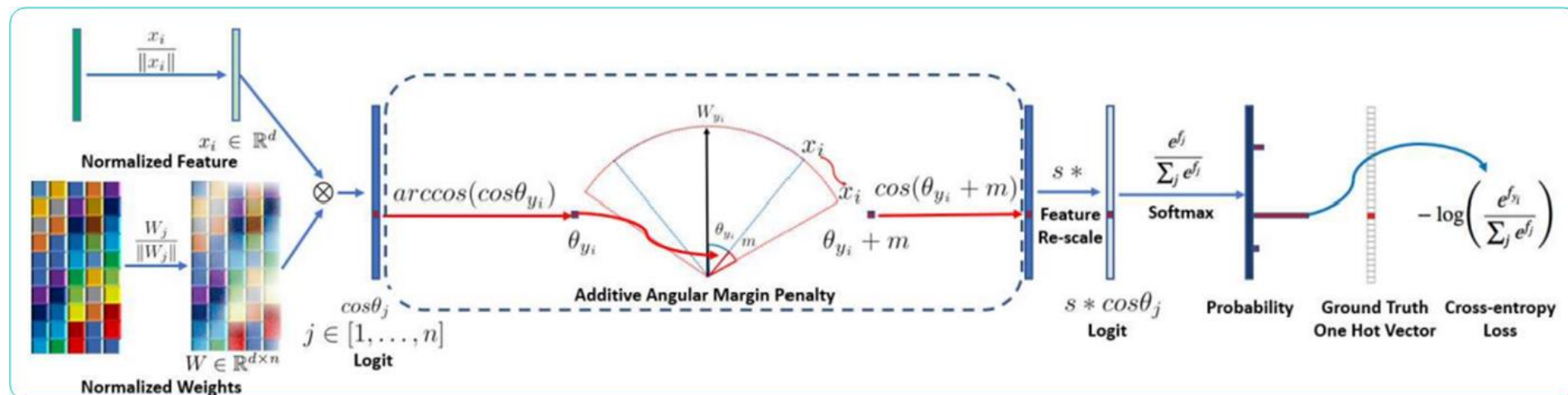
Верификация: «DeepFace»



- Classification network обучена на 4030 человек (класс) ~ 1000 изображений
- Постановка задачи: один vs другие

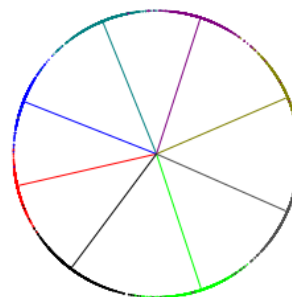


Идея нормализация и отступа

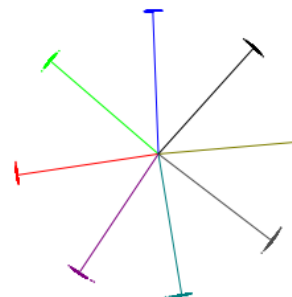


| Loss Functions | LFW | CFP-FP | AgeDB-30 |
|----------------------|--------------|--------------|--------------|
| ArcFace (0.4) | 99.53 | 95.41 | 94.98 |
| ArcFace (0.45) | 99.46 | 95.47 | 94.93 |
| ArcFace (0.5) | 99.53 | 95.56 | 95.15 |
| ArcFace (0.55) | 99.41 | 95.32 | 95.05 |
| SphereFace [18] | 99.42 | - | - |
| SphereFace (1.35) | 99.11 | 94.38 | 91.70 |
| CosFace [37] | 99.33 | - | - |
| CosFace (0.35) | 99.51 | 95.44 | 94.56 |
| CM1 (1, 0.3, 0.2) | 99.48 | 95.12 | 94.38 |
| CM2 (0.9, 0.4, 0.15) | 99.50 | 95.24 | 94.86 |
| Softmax | 99.08 | 94.39 | 92.33 |
| Norm-Softmax (NS) | 98.56 | 89.79 | 88.72 |

$$\text{ArcFace loss: } L_3 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}}$$

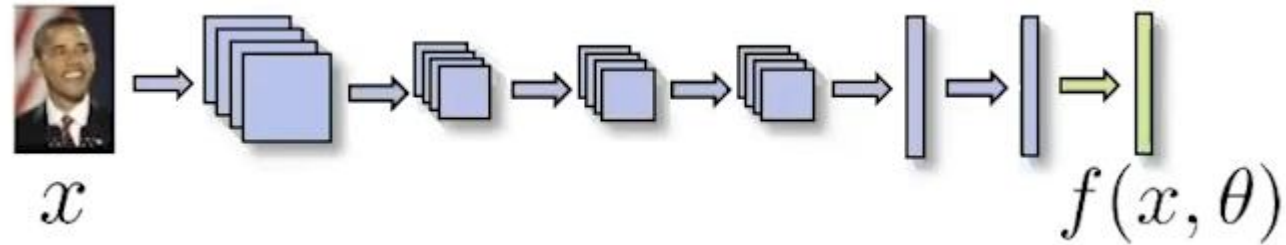


(a) Softmax



(b) ArcFace

Pair-based learning (contrastive)



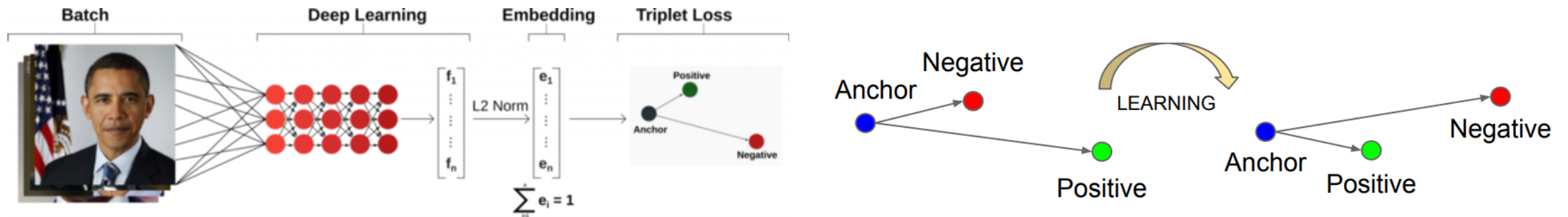
$$L^+((x_1, x_2); \theta) = \rho(f(x_1, \theta), f(x_2, \theta))$$

$$L^-((x_1, x_2); \theta) = \max(0, M - \rho(f(x_1, \theta), f(x_2, \theta)))$$

Функции расстояний:

- $1 - \cos$
- L2 (batch norm)
- Separate network

FaceNet



Примитивный **triplet loss**:
$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

- Использовать большие партии (mini-batches) – 1800, 40 изображений для разных классов + рандом
- Взять все positives для всей партии (batch)
- Взять «*semi-hard*» negatives:

$$\|f(x_i^a) - f(x_i^p)\|_2^2 < \|f(x_i^a) - f(x_i^n)\|_2^2$$

FaceNet: результаты



- Результаты 99.63% на LFW
(точность человека ~97%)

Точность и размер
выборка

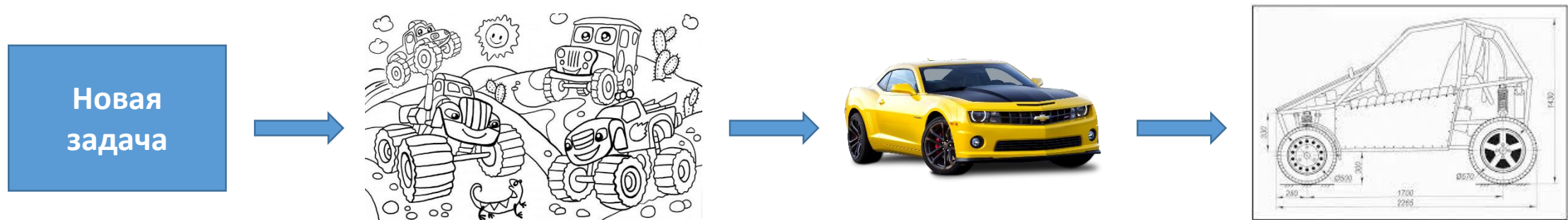
| #training images | VAL |
|------------------|-------|
| 2,600,000 | 76.3% |
| 26,000,000 | 85.1% |
| 52,000,000 | 85.1% |
| 260,000,000 | 86.2% |

Self-supervised feature learning

Общий подход:

- Взять модель, предобученную на текущих неразмеченных данных
- Дообучение (fine-tune) на новую задачу на размеченных данных

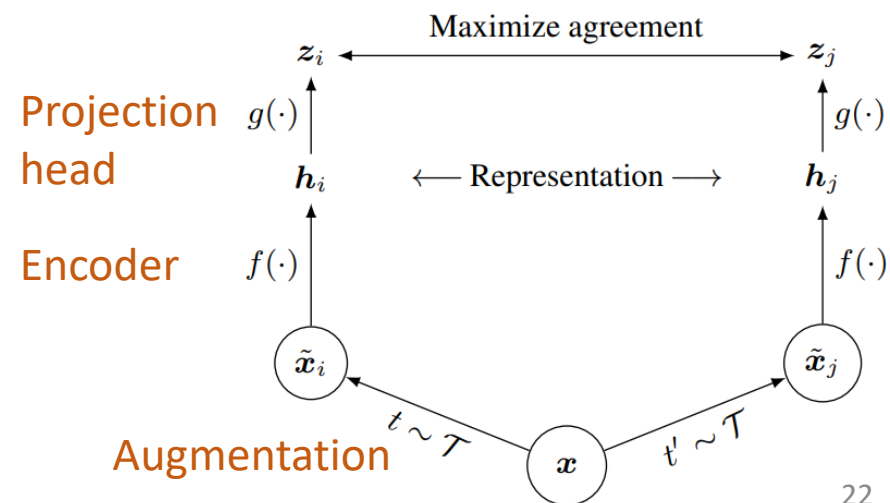
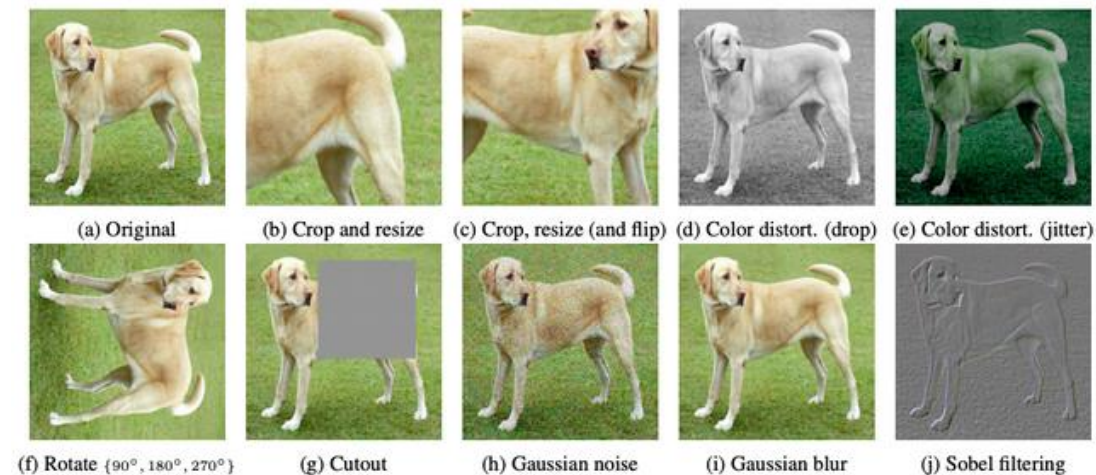
Альтернатива – взять предобученную модель на ImageNet



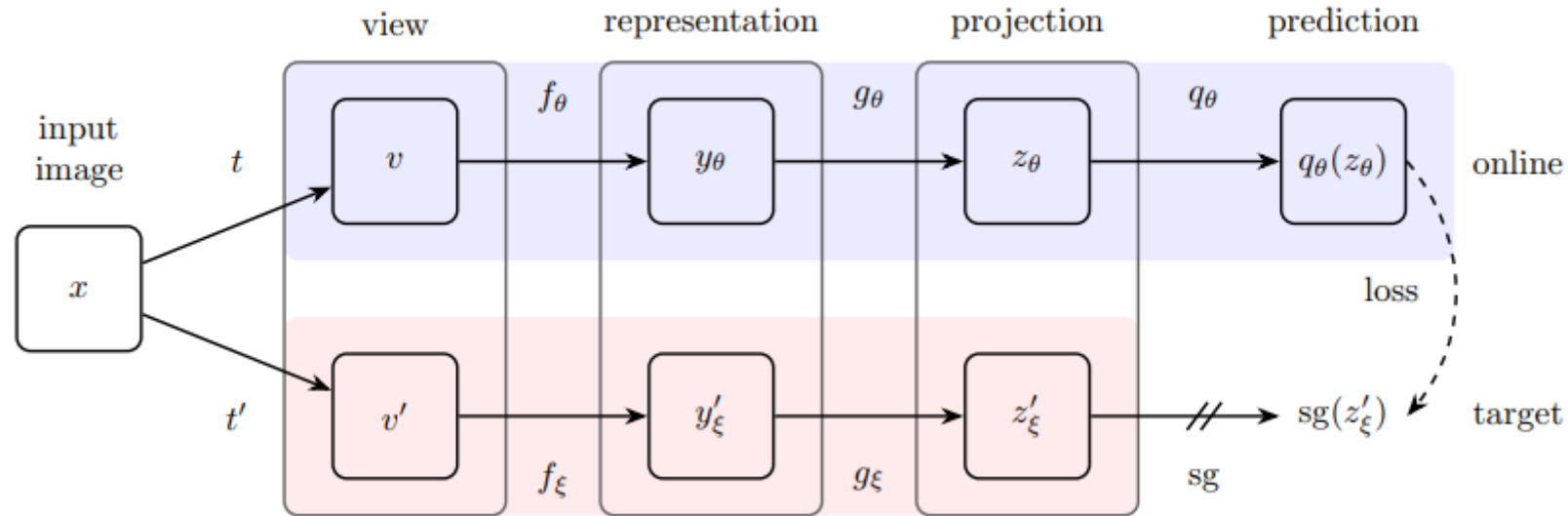
SimCLR

- Каждый batch содержит пары изображений
- Каждая пара является измененными версиями исходного изображения из выборки с изменениями типа crop, Gaussian blur, color distortion
- Цель – обучить модель сопоставлять пары, классифицировать как один объект (класс)
- Требуется большой размер пакета (batch)
- Loss:

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)}$$



Bootstrap Your Own Latent (BYOL)



- Расширение работы contrastive learning MoCo [He et al. CVPR 2020]
- Обучение без negatives
- SoTA для unsupervised pretraining

Transfer results

Классификация

| Method | Food101 | CIFAR10 | CIFAR100 | Birdsnap | SUN397 | Cars | Aircraft |
|---------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| <i>Linear evaluation:</i> | | | | | | | |
| BYOL (ours) | 75.3 | 91.3 | 78.4 | 57.2 | 62.2 | 67.8 | 60.6 |
| SimCLR (repro) | 72.8 | 90.5 | 74.4 | 42.4 | 60.6 | 49.3 | 49.8 |
| SimCLR [8] | 68.4 | 90.6 | 71.6 | 37.4 | 58.8 | 50.3 | 50.3 |
| Supervised-IN [8] | 72.3 | 93.6 | 78.3 | 53.7 | 61.9 | 66.7 | 61.0 |
| <i>Fine-tuned:</i> | | | | | | | |
| BYOL (ours) | 88.5 | 97.8 | 86.1 | 76.3 | 63.7 | 91.6 | 88.1 |
| SimCLR (repro) | 87.5 | 97.4 | 85.3 | 75.0 | 63.9 | 91.4 | 87.6 |
| SimCLR [8] | 88.2 | 97.7 | 85.9 | 75.9 | 63.5 | 91.3 | 88.1 |
| Supervised-IN [8] | 88.3 | 97.5 | 86.4 | 75.8 | 64.3 | 92.1 | 86.0 |
| Random init [8] | 86.9 | 95.9 | 80.2 | 76.1 | 53.6 | 91.4 | 85.9 |

Распознавание и сегментация

| Method | AP ₅₀ | mIoU |
|-------------------|------------------|-------------|
| Supervised-IN [9] | 74.4 | 74.4 |
| MoCo [9] | 74.9 | 72.5 |
| SimCLR (repro) | 75.2 | 75.2 |
| BYOL (ours) | 77.5 | 76.3 |

Применение

(1) Self-supervised learning on **unlabeled** natural images



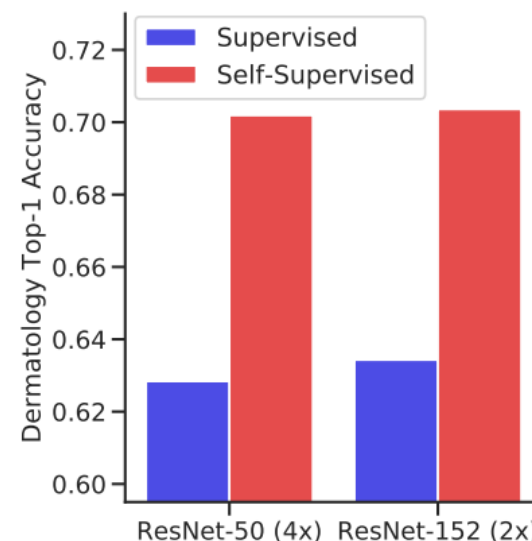
(2) Self-supervised learning on **unlabeled** medical images and **Multi-Instance Contrastive Learning (MICLe)** if multiple images of each medical condition are available



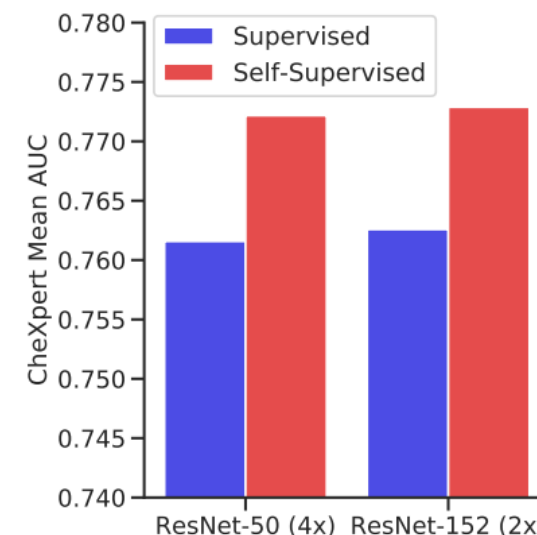
(3) Supervised fine-tuning on **labeled** medical images



15 000 обучающих
примеров,
454 000
неразмеченных,
27 классов



67 000 обучающих
примеров,
112 000
неразмеченных,
5 классов



Синее – предобучено на ImageNet

Красное – unsupervised ImageNet + Target domain