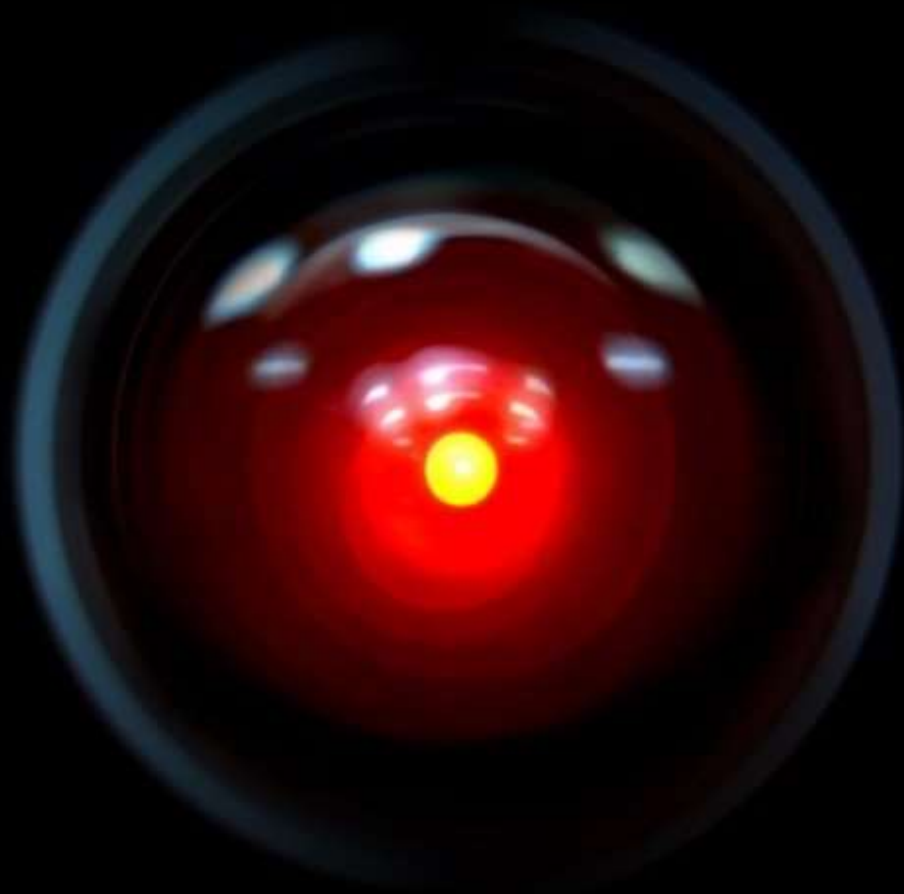


Как компьютерное зрение и машинное обучение помогают автоматизировать видеонаблюдение

Антон Конушин

Доцент НИУ ВШЭ, академический руководитель ОП ПМИ
Доцент ВМК МГУ, заведующий лабораторий компьютерной
графики и мультимедиа ВМК МГУ
Лектор Школы Анализа Данных Яндекса

9 декабря 2017 года, Санкт-Петербург, ИТМО





Распознавание лиц по видео с камер



Самый популярный элемент «искусственного интеллекта» в кино и на ТВ



Сколько всего камер?





You have been watched





Социальный рейтинг



Фантастика в BlackMirror: Nosedive



Социальный рейтинг: jaywalking



Реальный эксперимент и глобальные планы в Китае



SenseTime



- SenseTime - один из «стартапов» smart surveillance с \$410M инвестициями
- Пилотные проекты с полицией и большими успехами
- 2017 – полное видеопокрытие Пекина

Другой пример – определение пьяных в метро. 230 смертей в Японии от поездов, 60% из них были пьяны.



Распознавание на 120 км/час





А что у нас?



**Инфогород. Городская
система видеонаблюдения**

Телефон Общегородского Контакт-центра:
(495) 587-00-02

[О системе](#) | [Окно в город](#) | [Проекты](#) | [Документы](#) | [Параметры камер](#) | [Обзор камер](#) | [Вопросы-ответы](#) | [Контакты](#)

**Городская система
видеонаблюдения** —
это государственная информационная
система для сбора, обработки и
хранения видеозаписей с камер,
установленных в городе,
предназначенная для повышения
качества жизни населения города
Москвы и уровня обеспечения
безопасности.

[Узнать больше](#)





Размеры и покрытие



Дворовое видеонаблюдение



Видеонаблюдение в местах
массового скопления
граждан



Подъездное
видеонаблюдение

На сегодняшний день камерами оборудовано более 100 тыс. подъездов, около 20 тыс. дворовых территорий и более 2,5 тыс. общественных мест.



Камеры рядом с ФКН НИУ ВШЭ



ПОРТАЛ ОТКРЫТЫХ ДАННЫХ

Правительства Москвы

ДАННЫЕ

СПРАВОЧНИКИ

ПРИЛОЖЕНИЯ

НОВОСТИ

ИНФОРМАЦИЯ

ФОРУМ

КАРТА

ОБРАТНАЯ СВЯЗЬ

СВОИ ПРОДУКТЫ!

Разработчикам

ENG



Реестр камер дворового видеонаблюдения

Просмотров: 1036116

Скачиваний: 10012

Оценка полезности:

2.5

RSS

Таблица

Карта

Паспорт

Описание

Скачать



Не нашли объект?

Строка №1008

Сообщить о некорректных данных

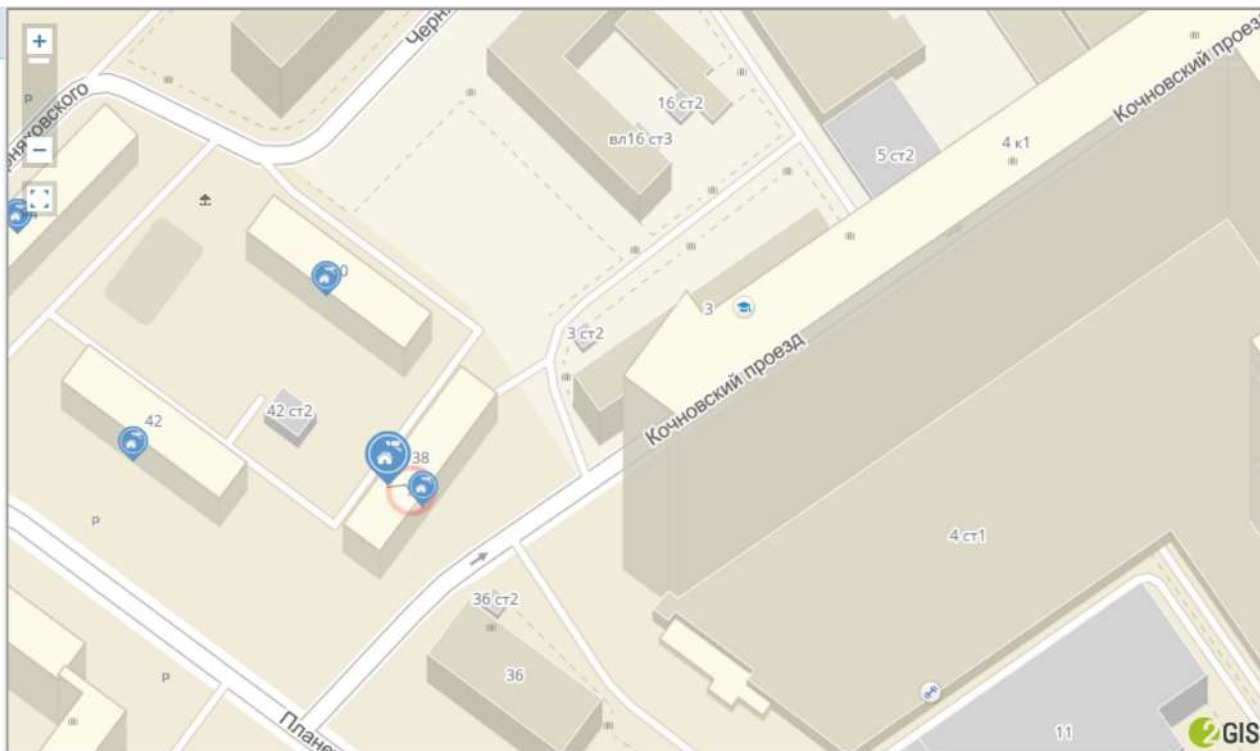
Район:
район Аэропорт

Код:
DVN_SAO_2_6695_1

Адрес ОВД:
улица Черняховского, дом 10

Телефон ОВД:

Фотография:



Упс...



Как этими данными воспользоваться?



Случилось
происшествие?



Узнайте, установлены ли
камеры на data.mos.ru



Оставьте заявку на сохранение
видеоархива по телефону
+7 (495) 587-00-02



Обратитесь в полицию



Полиция получит видеоархив
и приложит к делу



Виновник наказан —
вы довольны

А вообще хороший и открытый вопрос....



Компании в России

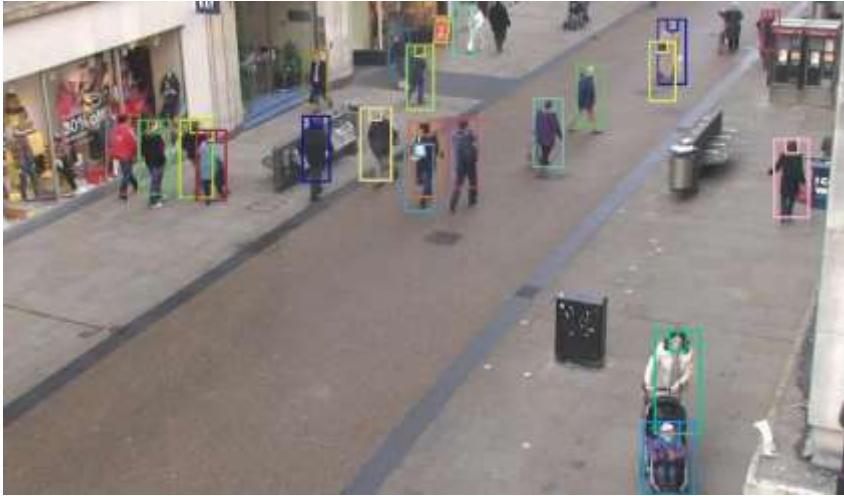


NTechLabs

- VOCORD
- 3DiVi
- Центр Речевых Технологий
- etc.



Что нужно извлекать?



- Выделить все объекты и людей в кадре
- Отследить все объекты и людей в видео
- Идентифицировать все объекты и людей
- Распознать происходящие события
- Занести и проиндексировать в архив
- Спрогнозировать развитие ситуации



Компьютерное зрение



Задача зрения: понять, что находится на изображении

Тест Тьюринга для компьютерного зрения:

Ответить на любой вопрос про изображении, на который может ответить человек.

Компьютерное зрение – часть области искусственного интеллекта (AI)



Что и где находится на изображении?

Выделение объектов

Здание

Текст

Лицо

Текст

中华人民共和国万岁

世界人民大团结万岁

Человек

Человек

Лицо

Необходимо определить, есть ли на изображении объекты заданного типа и если да, то определить их положение



Какой?

Ветер
слабый

Голубое

Наклонная

Мао

Анфас

Профиль

Характеристики (атрибуты) отдельных объектов



Почему это сложно?



Какие свойства изображений объектов делают задачу сложной?



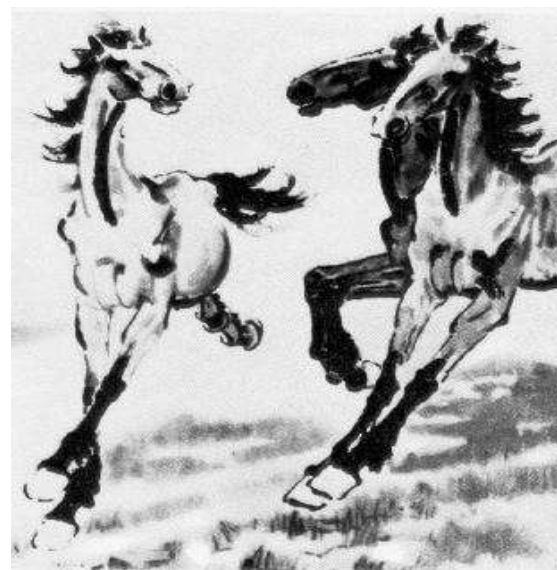
Почему это сложно?



Ракурс съёмки



Освещение



Внутриклассовая изменчивость, размер, деформация объектов



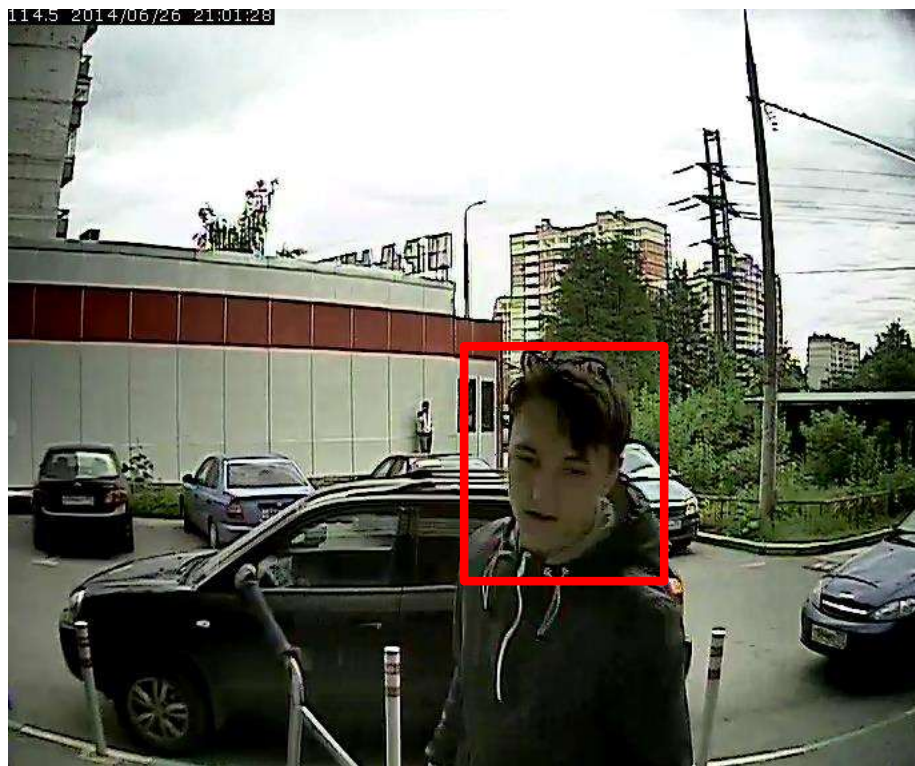
Применительно к видеонаблюдению



- Ракурс, вид наблюдаемых объектов и т.д.
- Ракурсы съёмки могут быть крайне различны
- Работающие системы удастся создать, «заточившись» на определённый сценарий съёмки



Что нам нужно научиться делать?



Находить людей и их лица
(детектирование, выделение
лиц)



Идентифицировать
(кто это?)

Как нам подойти к решению этой задачи?



Сведение к классификации

Задача классификации – определить для объекта x его «класс» y из заданного конечного набора классов

Объект x_1



$y_1 = 1$ (да)

Объект x_2



$y_2 = 0$ (нет)

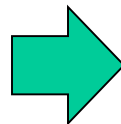
Это лицо?

Классификатор $f(x) = y$ – функция, отображающая x в соответствующий класс



Выделение объектов на изображении

Задача поиска объектов сводится к классификации



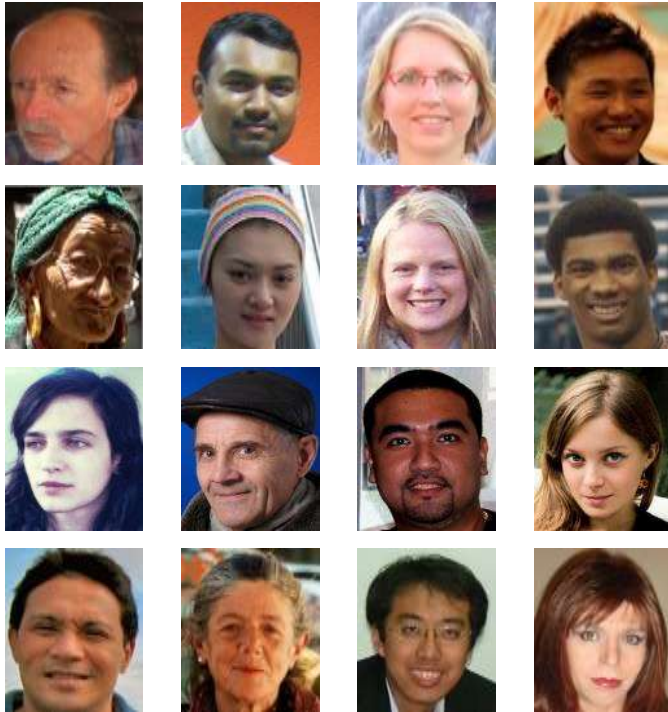
Разделили изображение на фрагменты, которые классифицируем независимо друг от друга

- «Скользящее окно» – сканирование окном изображения
- Каждый фрагмент отдельно от других классифицируем «объект» / «не объект»



Что такое «лицо»?

- Соберём много примеров «лиц» и «не лиц»



Лица

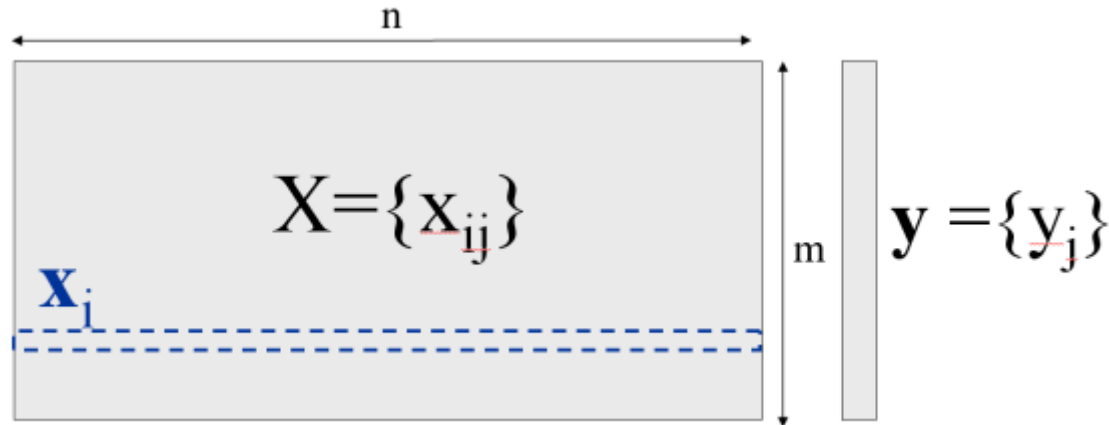


Не лица



Построение классификатора

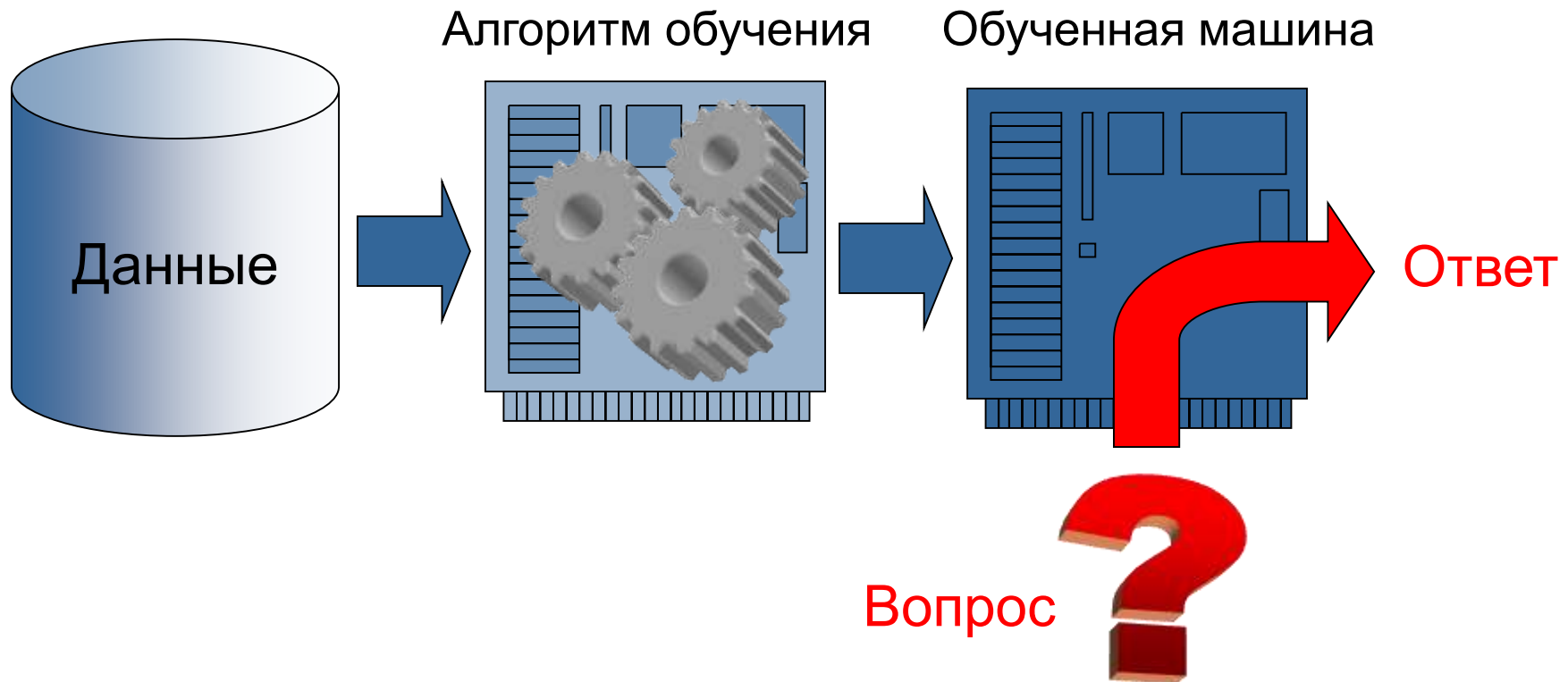
- Есть выборка данных X для которых известны Y
- Каждый объект из X с номером j можно описать вектором признаков \mathbf{x}_j (набором характеристик)
- Всё множество известных наблюдений (конечное) можно записать в следующем виде (обучающая выборка):



Нужно построить функцию $f(x) = y$



Как должно быть в идеале

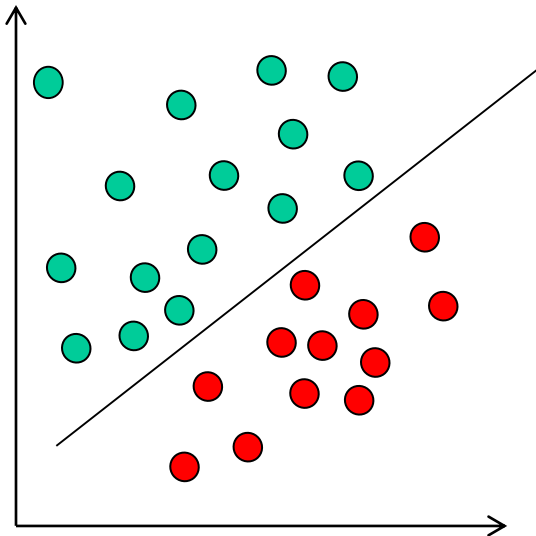
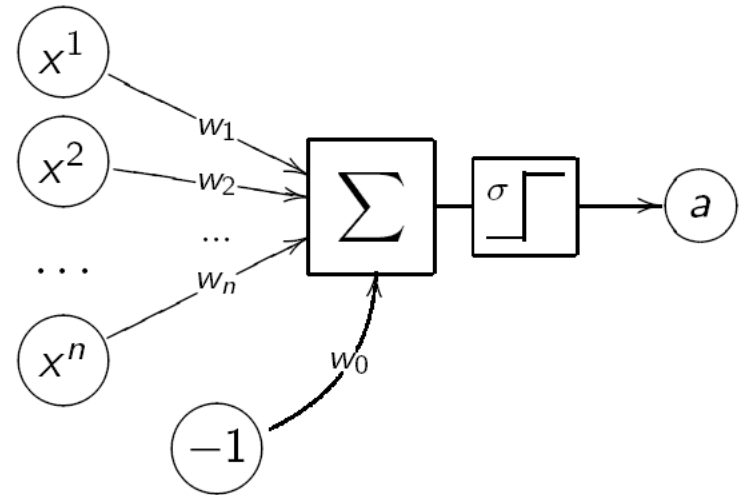
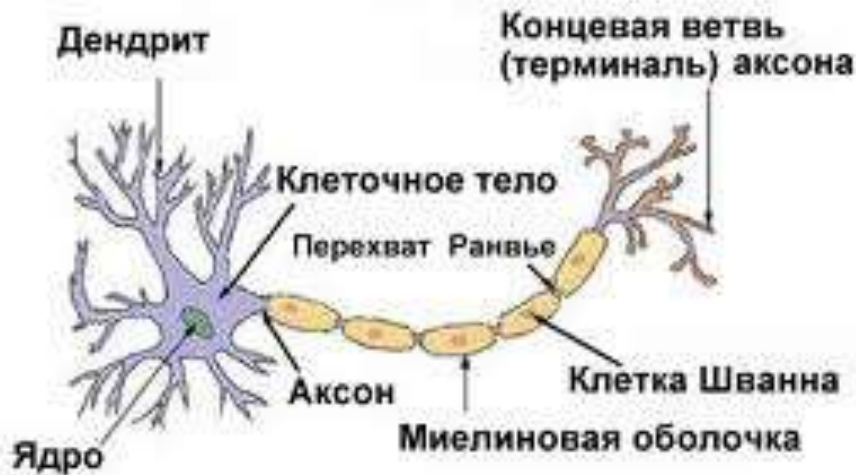


Машинное обучение



Нейрон и нейронная сеть

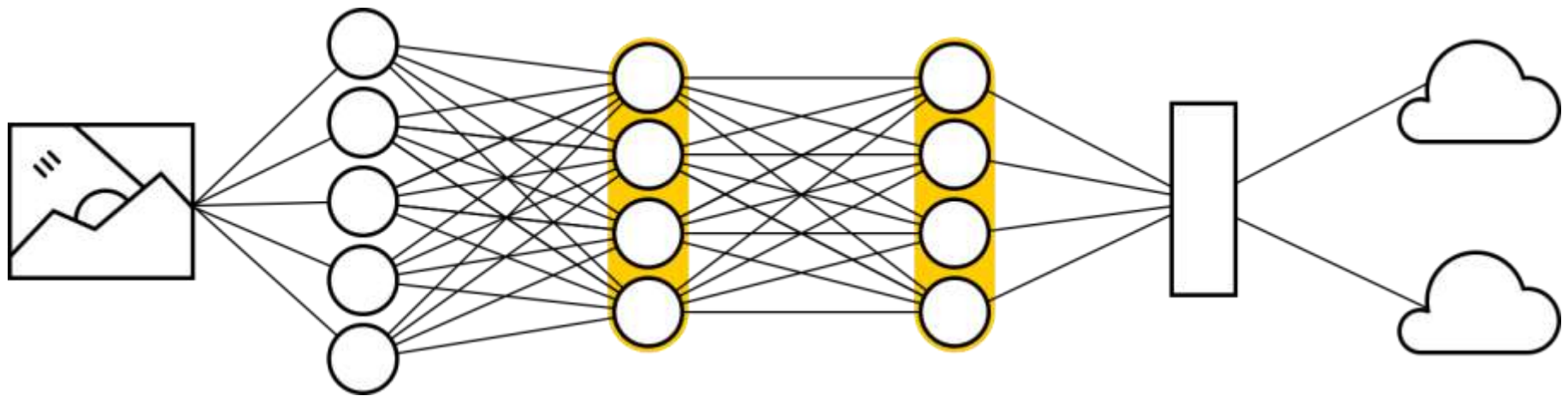
Типичная структура нейрона



- Нейрон как линейный классификатор
- Разбивает пространство объектов на две области плоскостью



Нейросеть



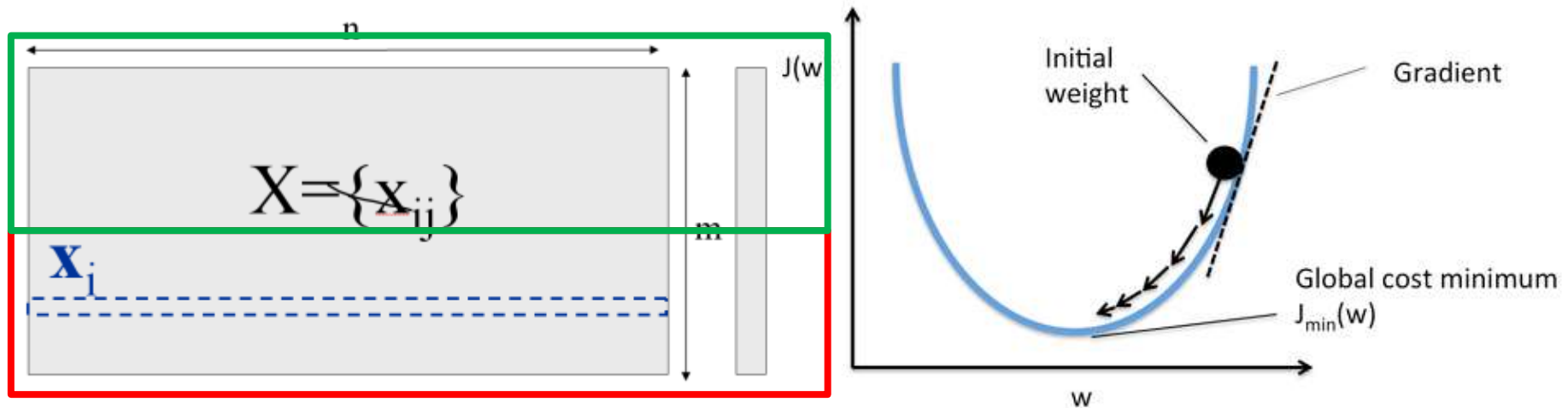
Архитектура – структура графа (связи нейронов)

Параметры сети – совокупность весов каждого нейрона



Как обучаем?

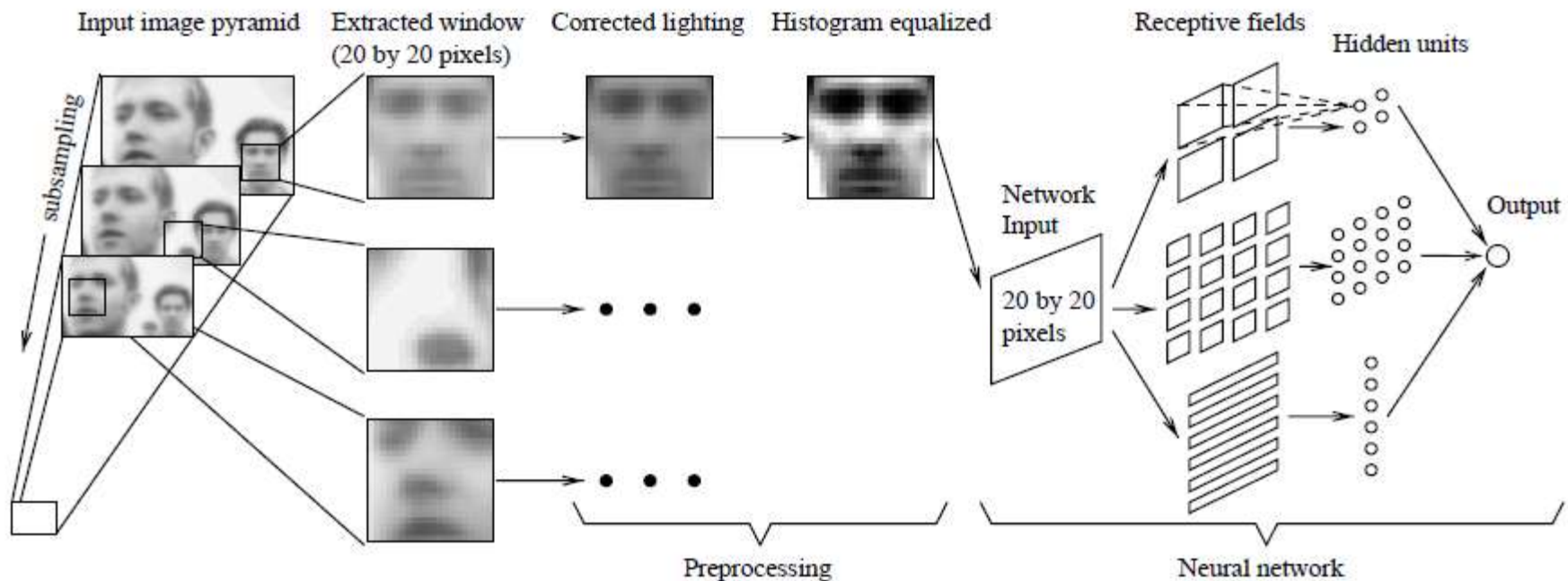
- Берём имеющиеся данные (тренировочную выборку)
- Определим функцию качества решения задачи, называемую функцией потерь (loss)
- Применяем метод оптимизации, и ищем оптимальные параметры (веса нейросети), которая дает минимум функции потери





Rowley face detector (1998)

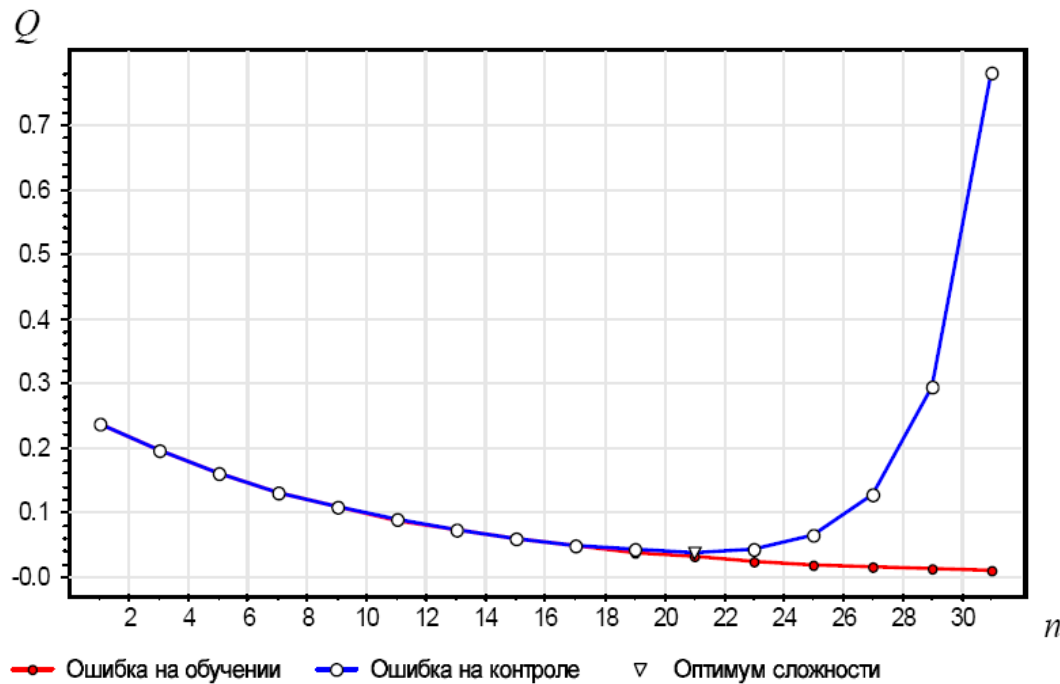
- Стохастически градиентный спуск, он же метод обратного распространения ошибки оказался (70-80-90е) очень эффективным
- Пример – детектор лица, лучший до Viola-Jones



B. Rowley, T. Kanade. Neural Network-Based Face Detection. PAMI, 1998.



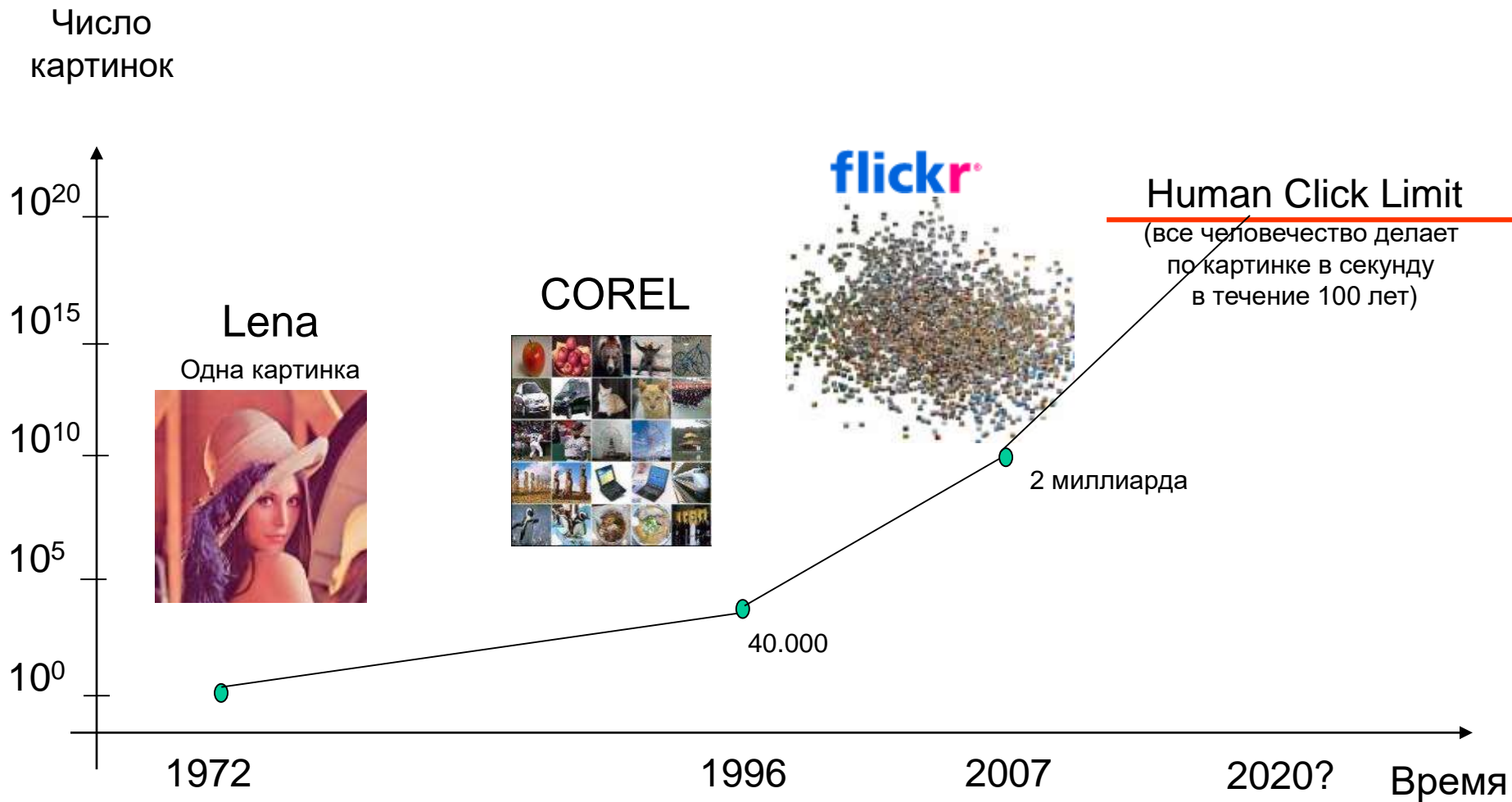
Переобучение



- Чем сложнее задача – тем более сложная нейросеть нужна
- Но параметрой нейросети очень много, и нейросеть быстро «переобучалась»
- Происходило «запоминание» всей обучающей выборки без её «обобщения»

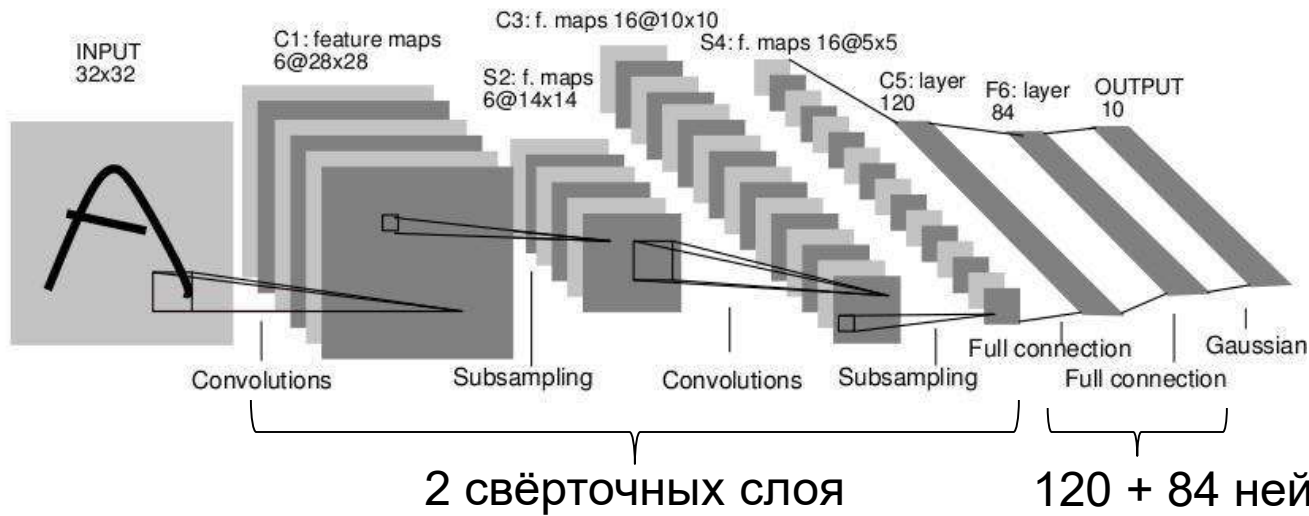


«Интернет-бум» + «Закон Мура»

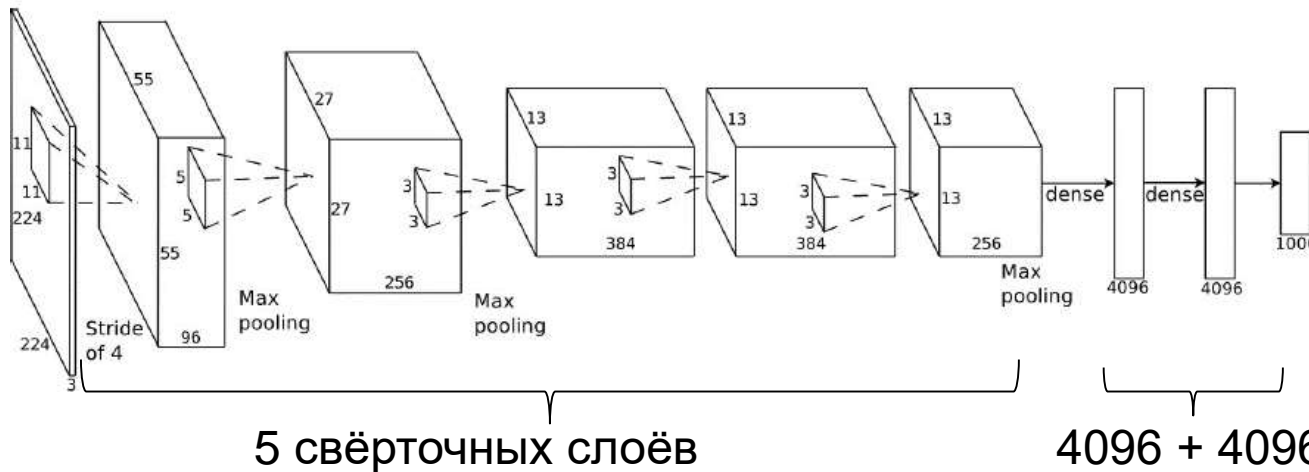




Свёрточные сети 1998 - 2012



LeNet
1998



AlexNet
2012

Всё учим на одном компьютере!



Развитие нейросетей



1998



2012



2014



2014



02.2015



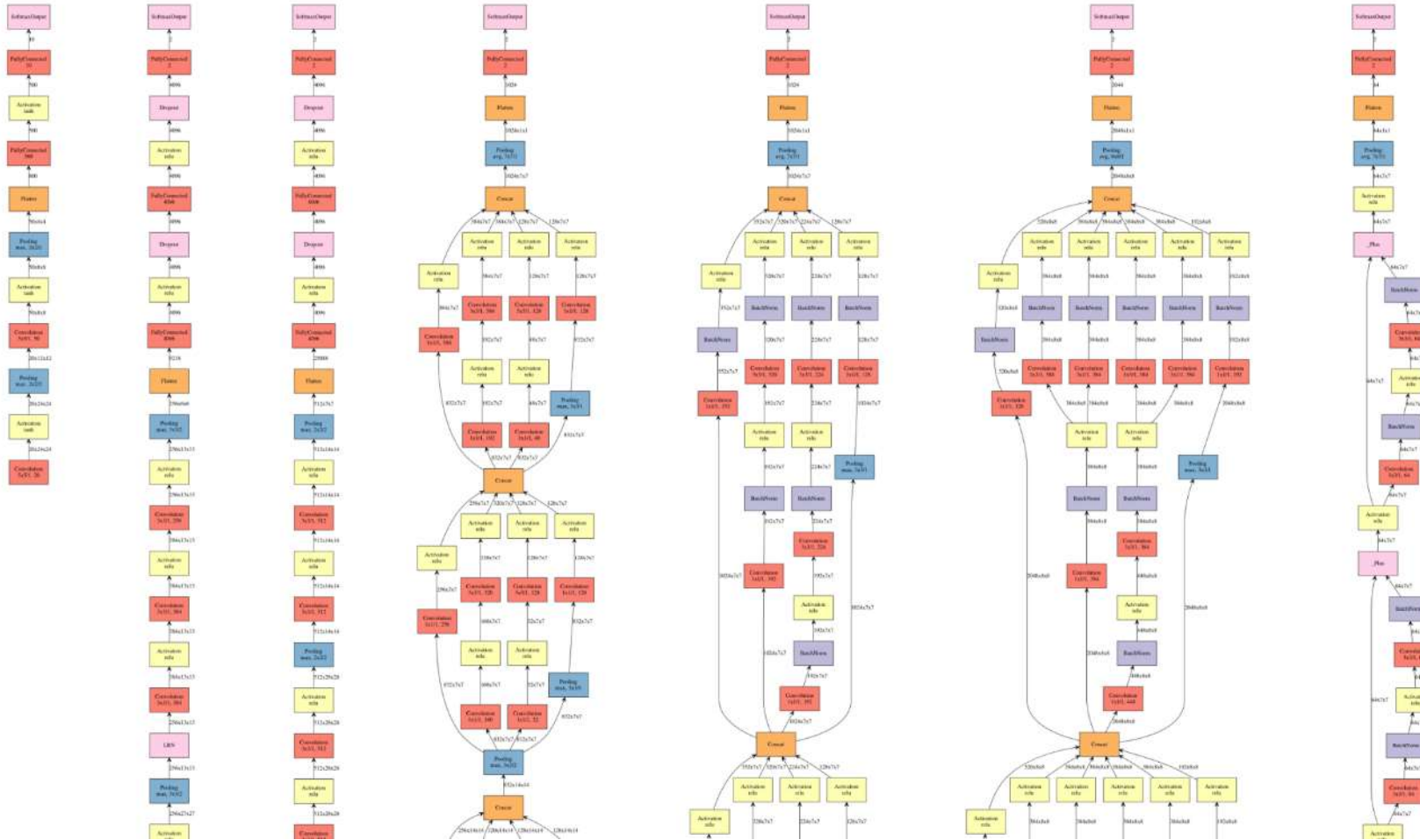
12.2015



12.2015

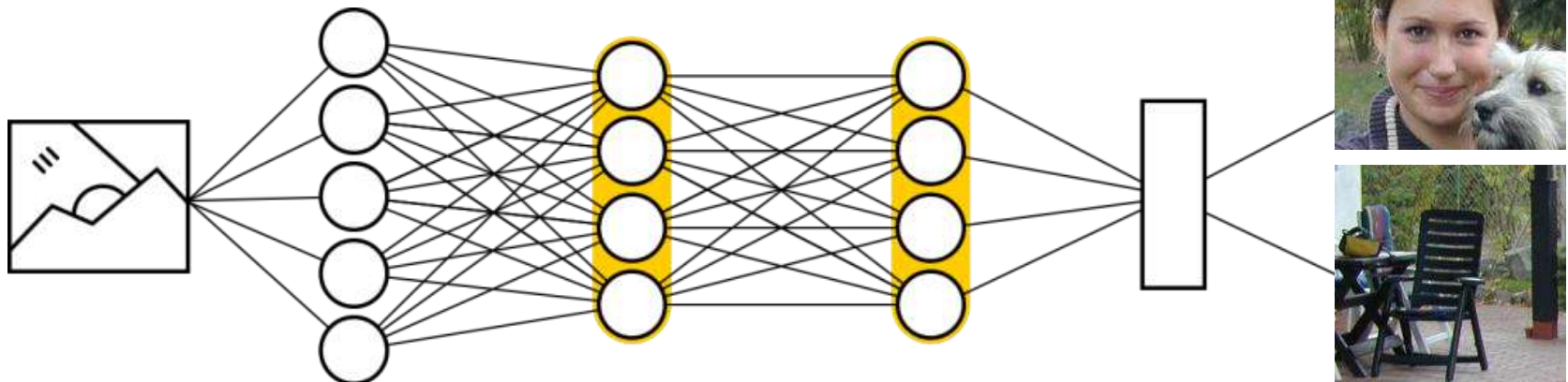
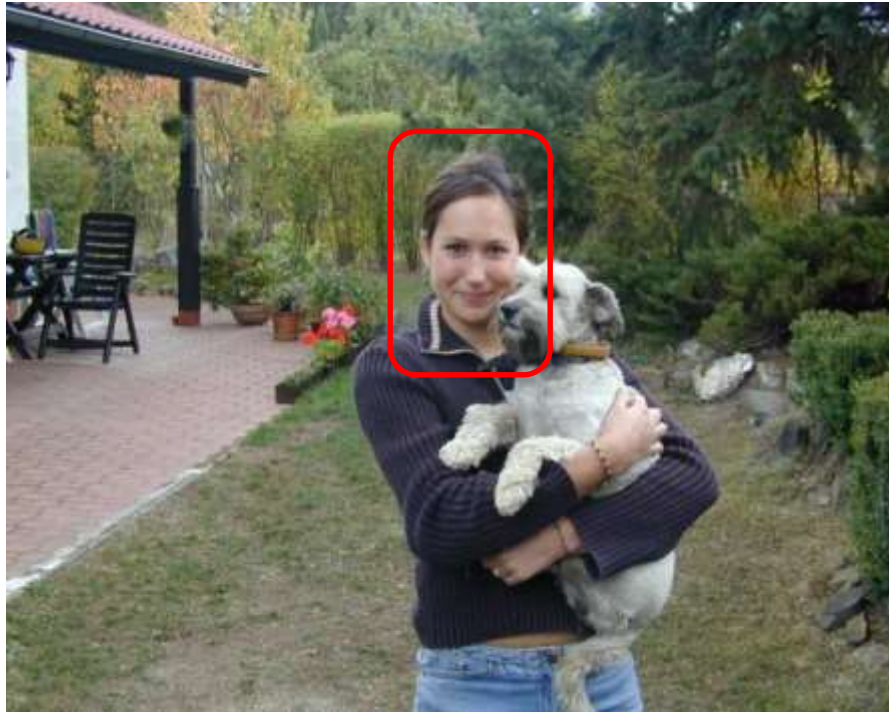


Развитие нейросетей



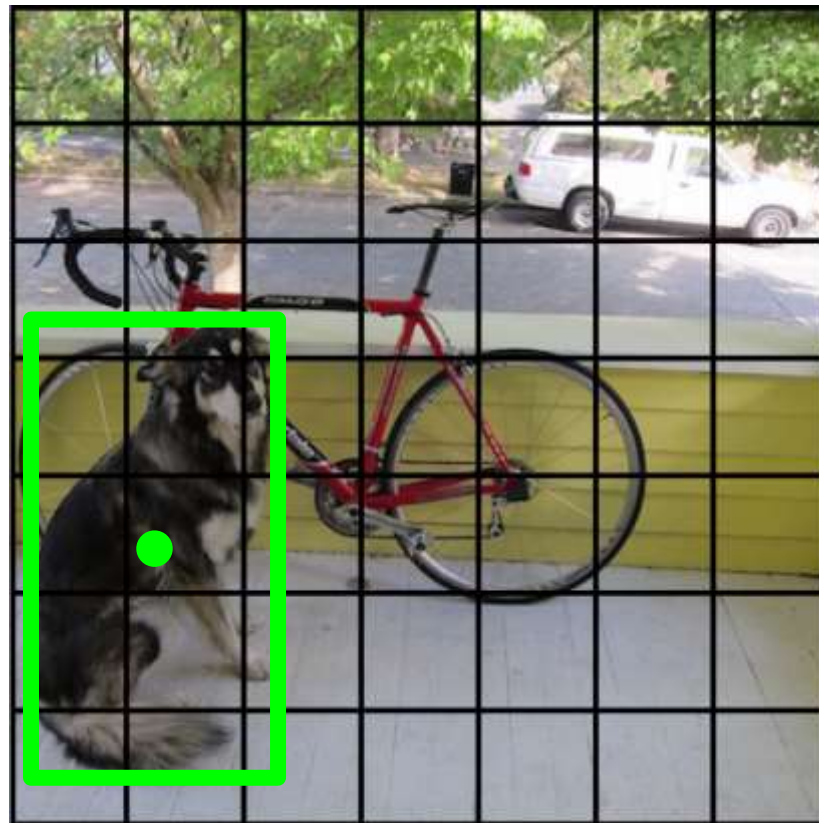
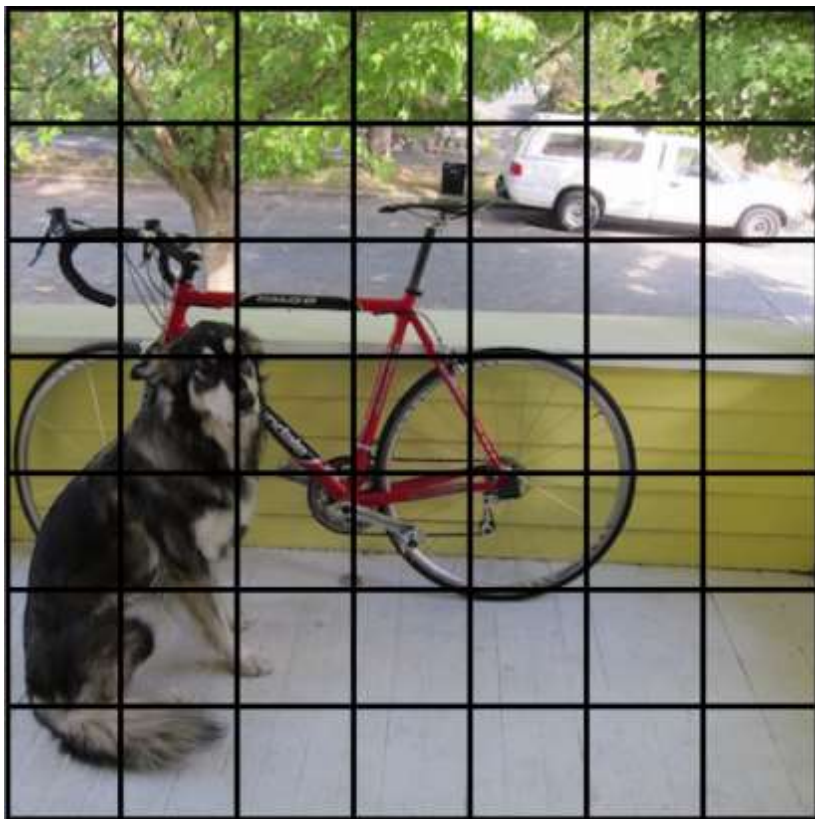


Вернёмся к детектированию лиц



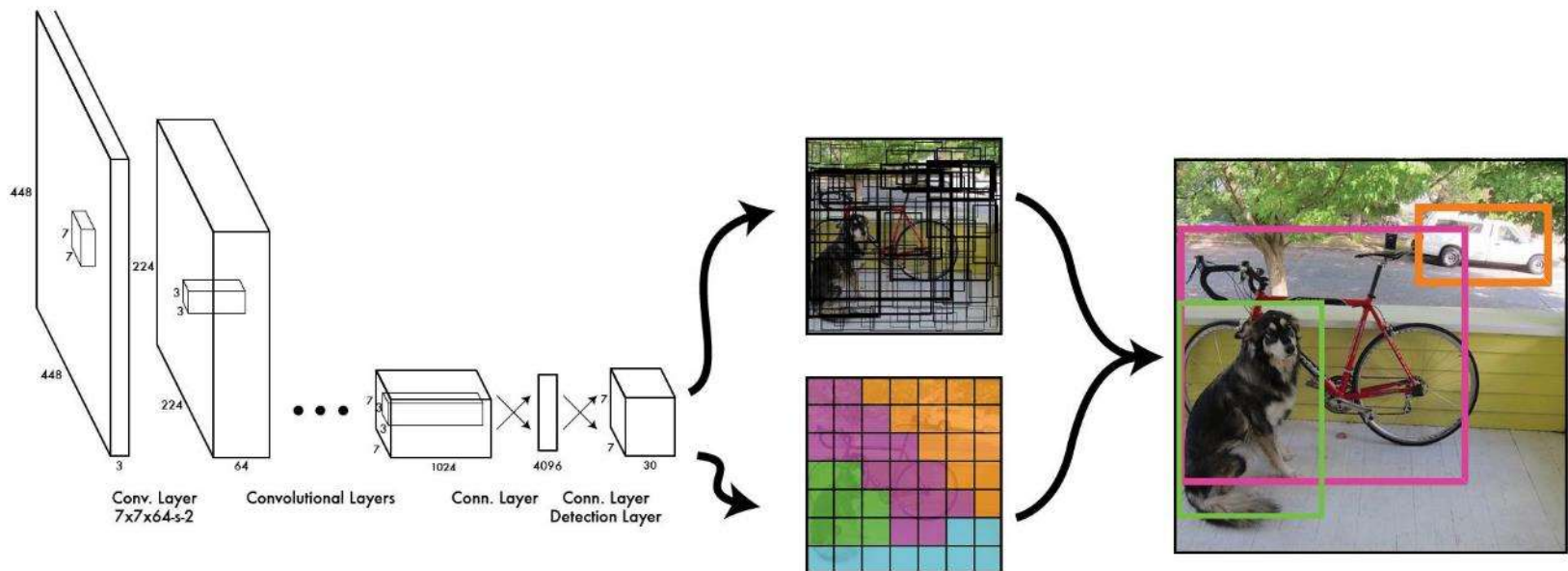
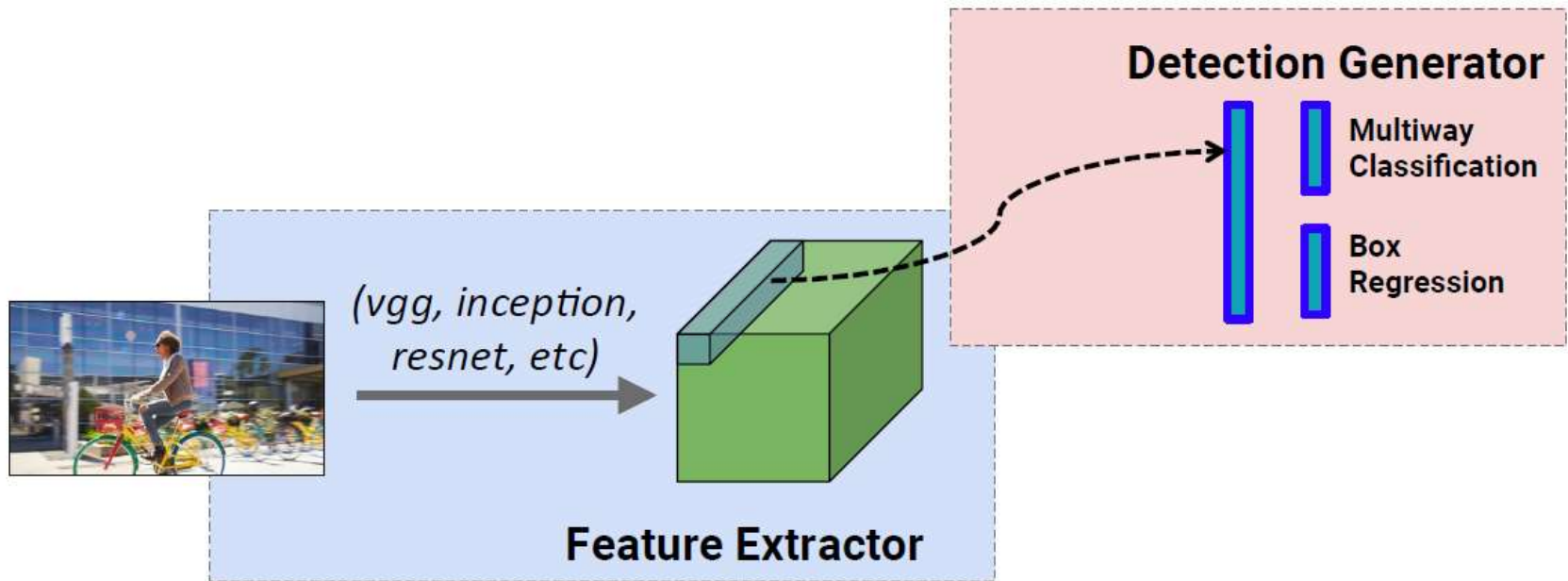


Регулярная сетка вместо окна





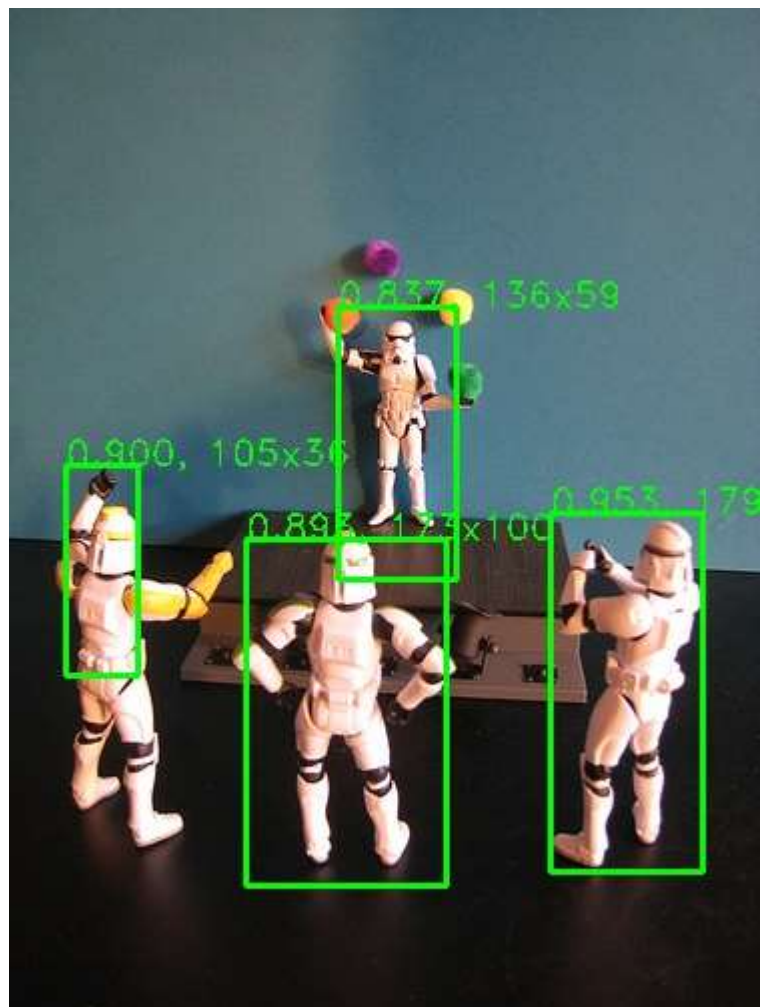
Single Shot Detector





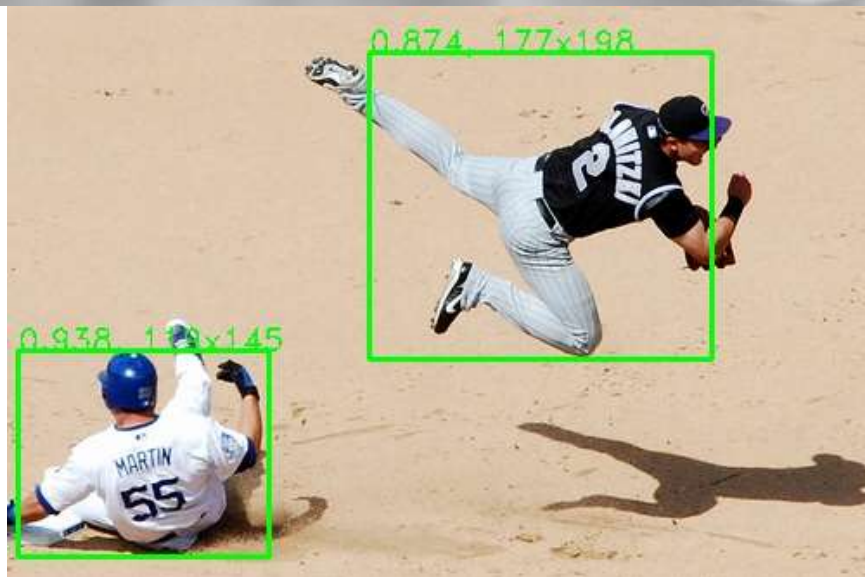
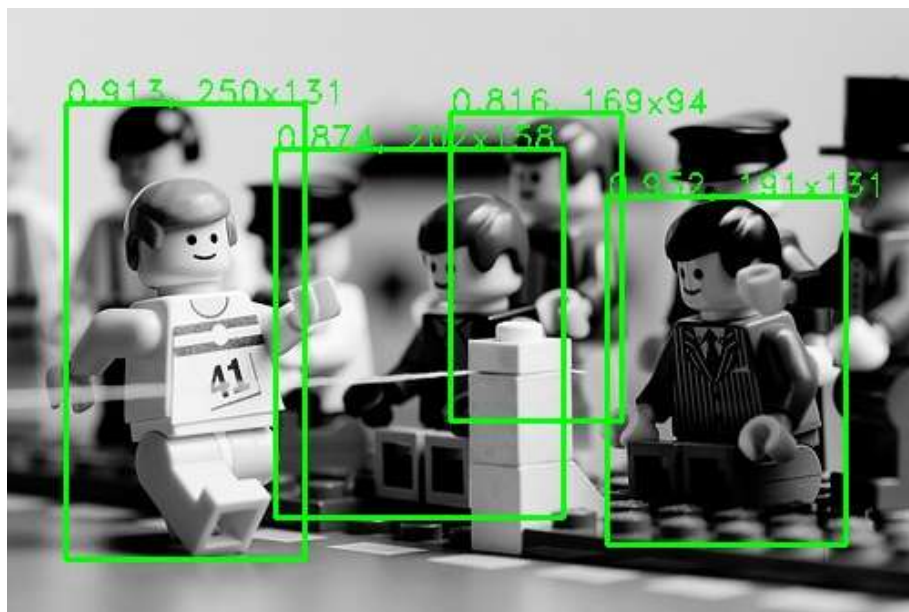


Причуды детекторов пешеходов



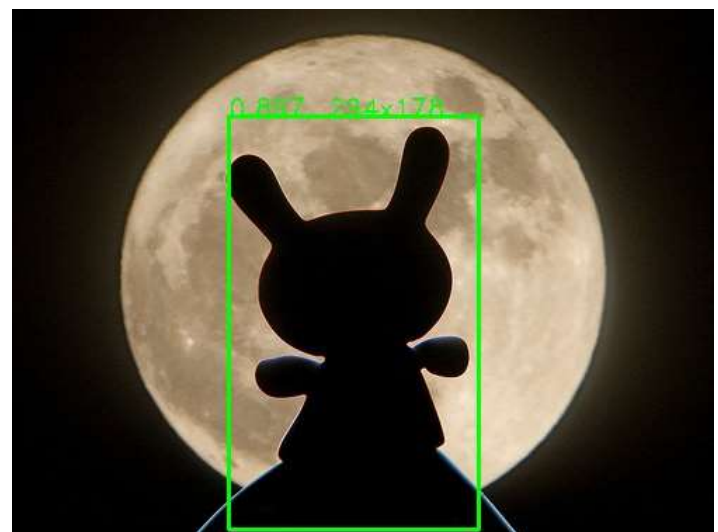
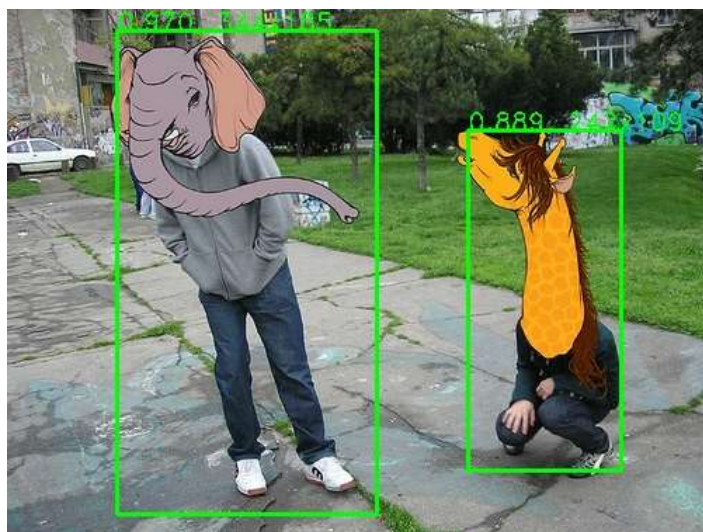
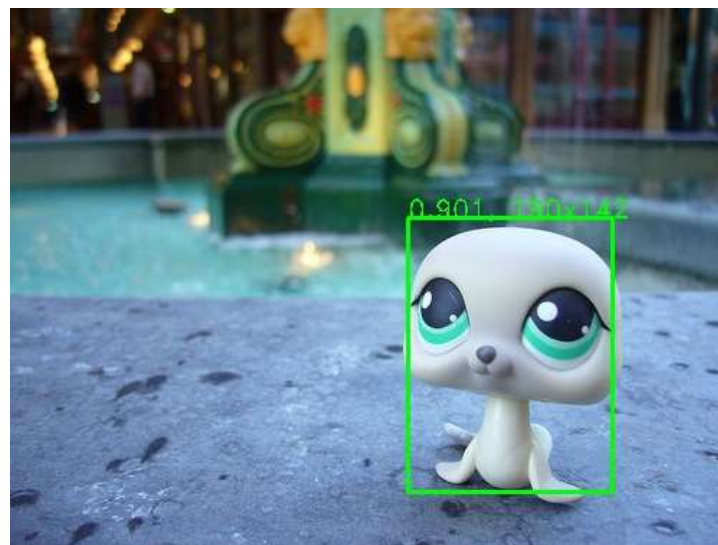
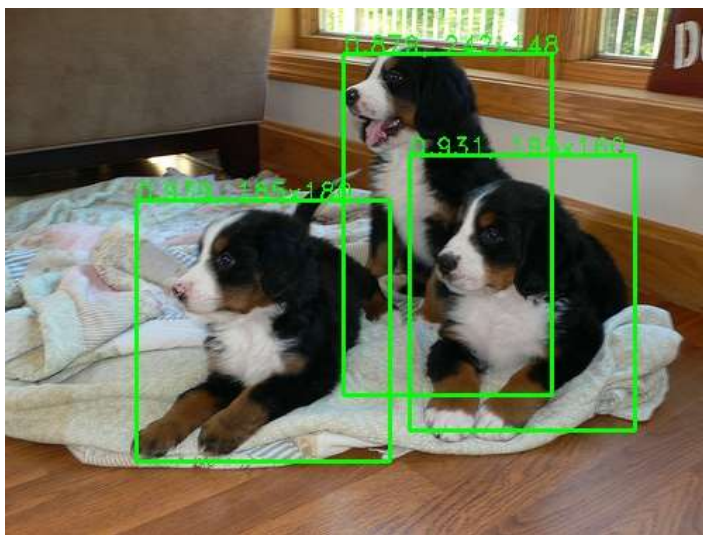


Причуды детекторов



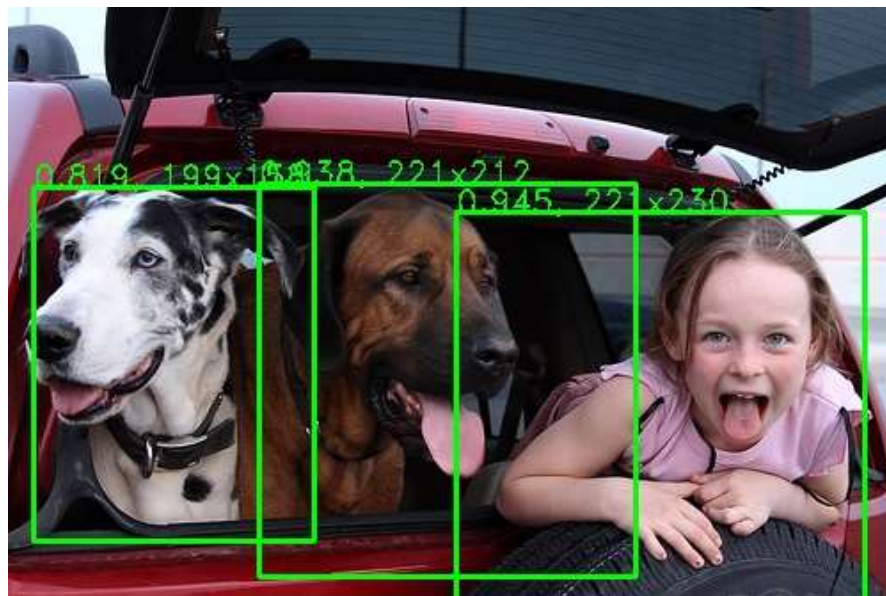
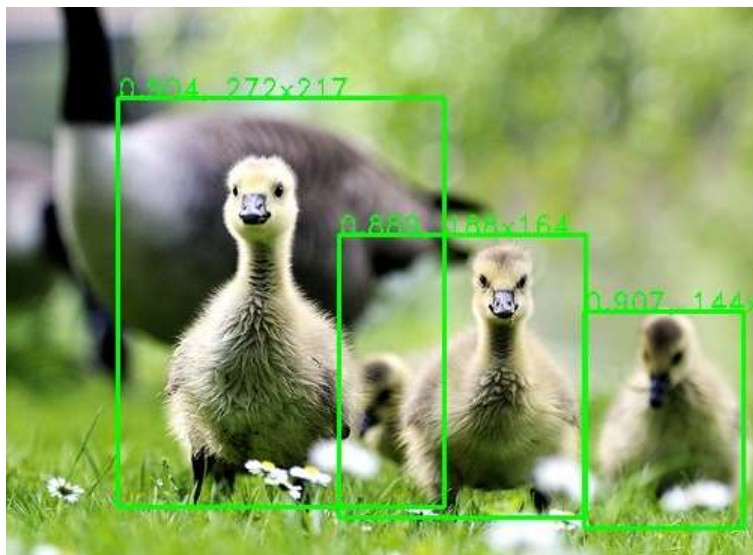


Причуды детекторов пешеходов



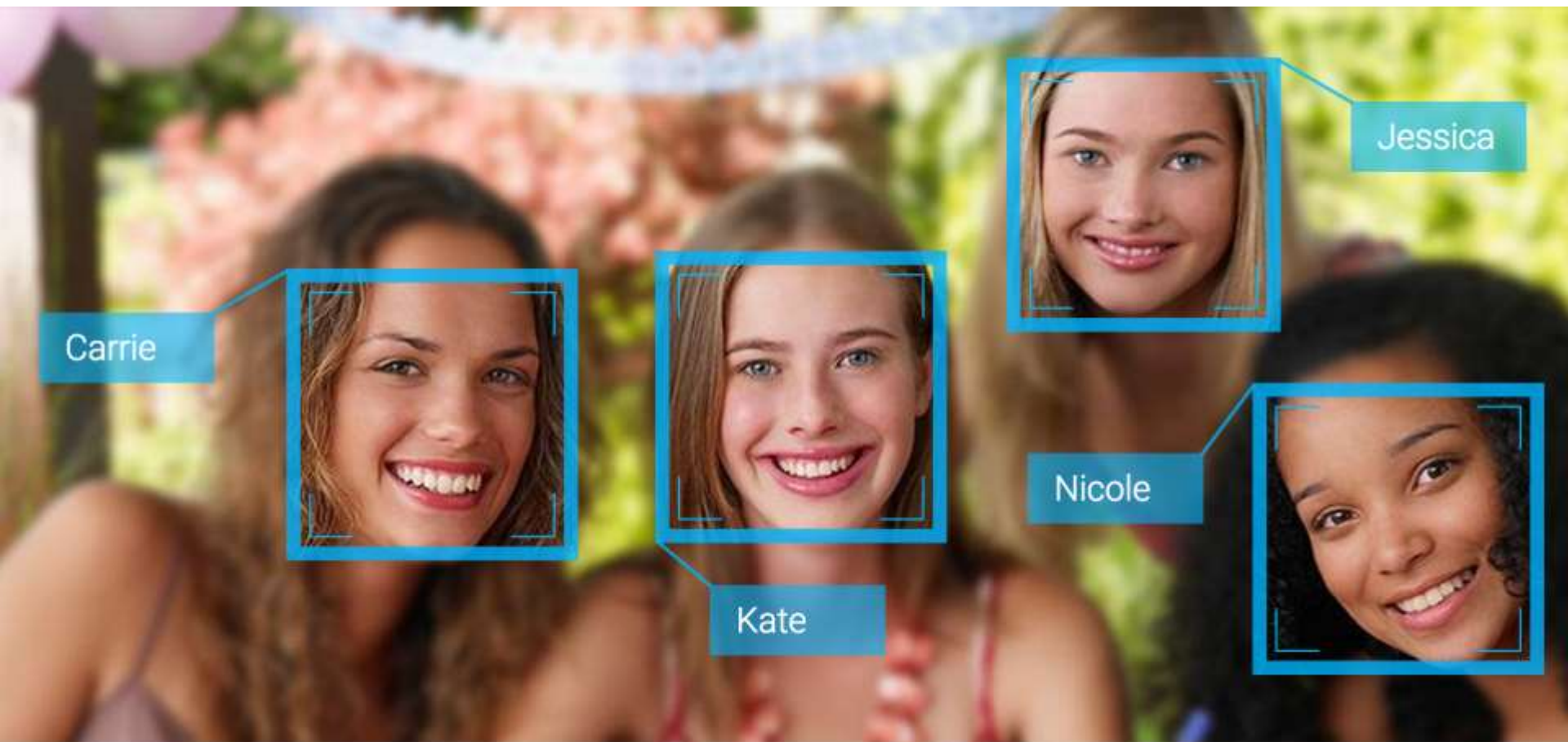


Причуды детекторов пешеходов





Распознавание лиц



Carrie

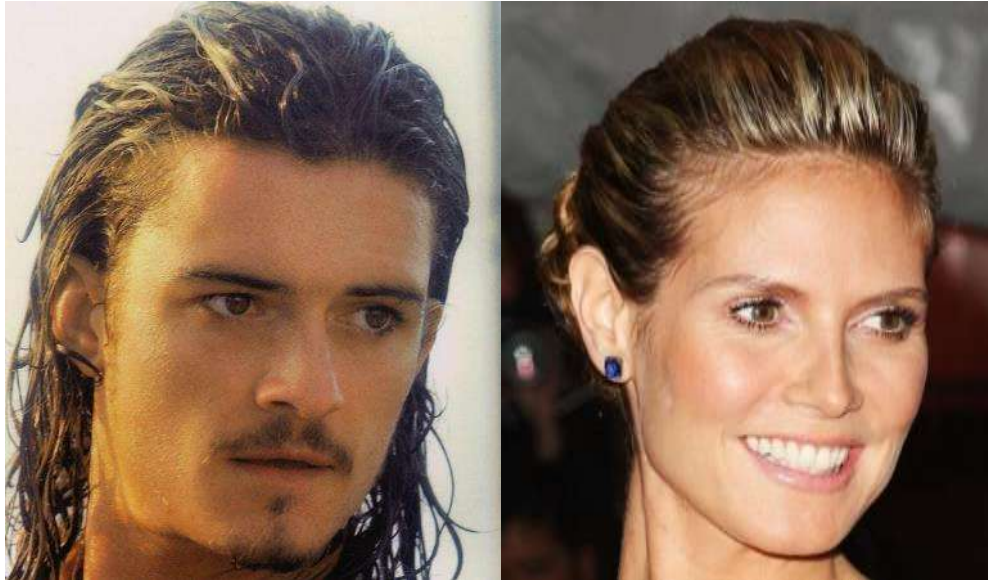
Kate

Jessica

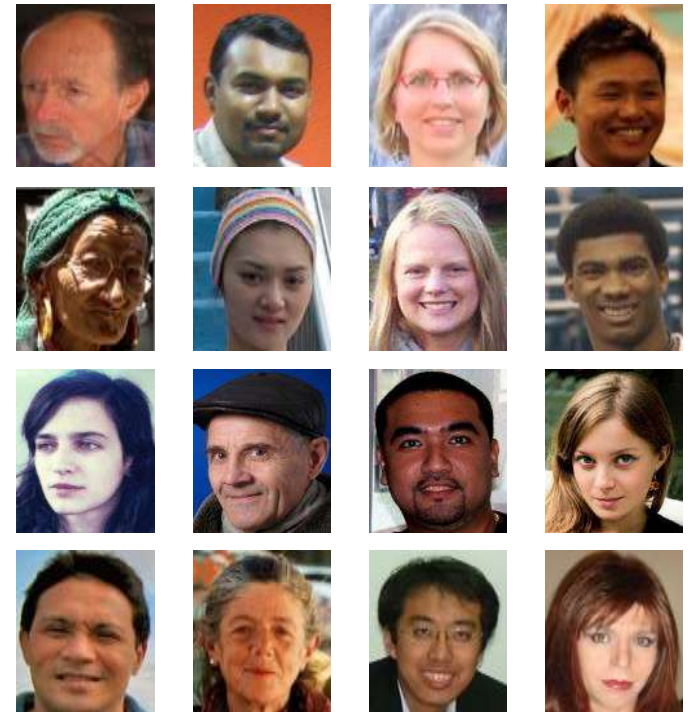
Nicole



Постановка задачи



Один и тот же человек ($y=1$),
или разные ($y=0$)?
(бинарная классификация)
«Верификация»



Кто из списка?
«Идентификация»

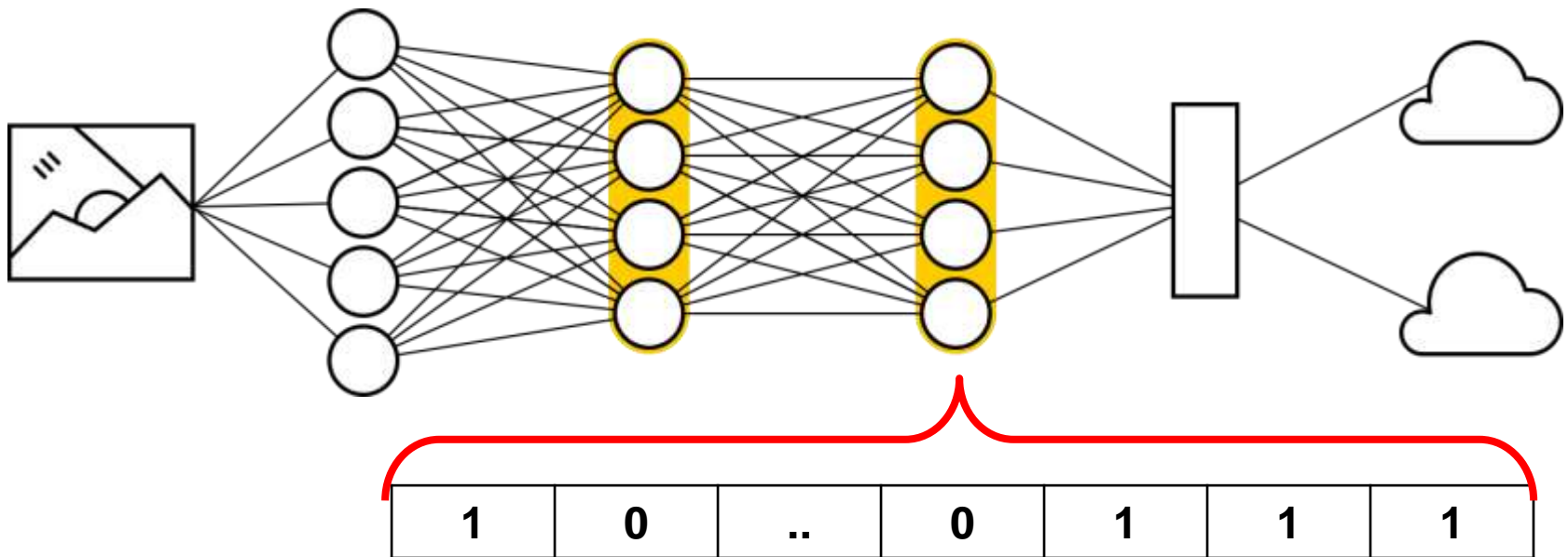


Сложность реальной жизни



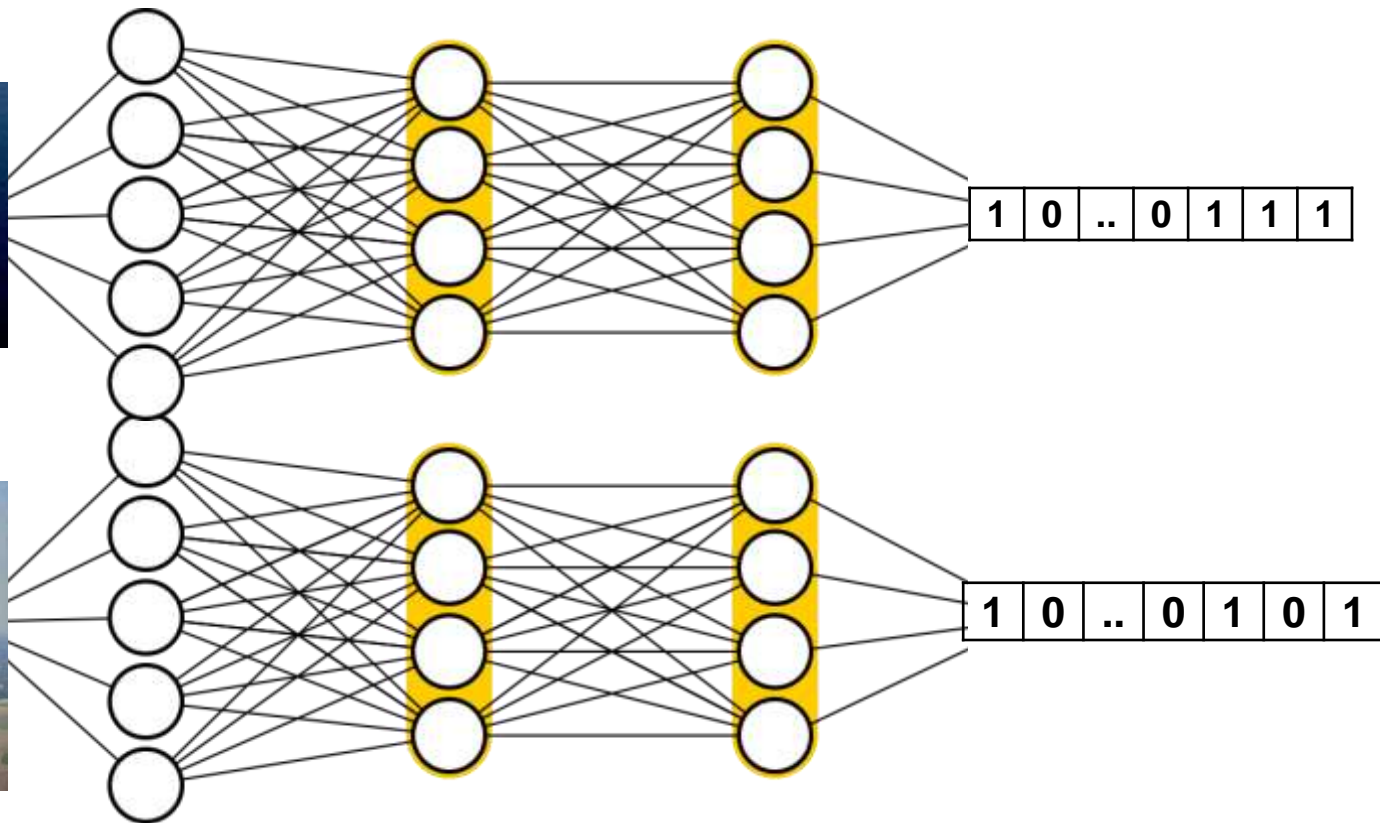


Вектор-признак



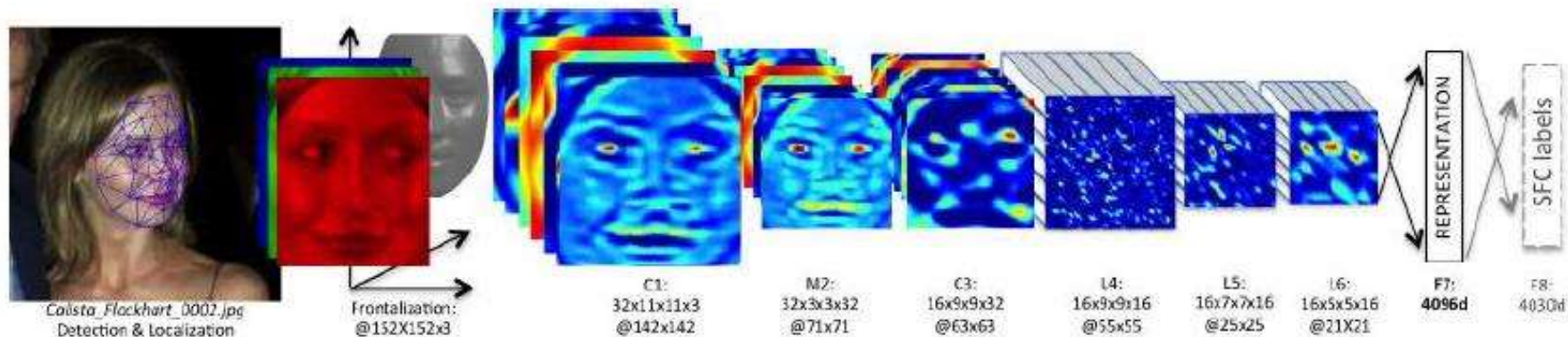


Сравнение изображений





DeepFace (Facebook)



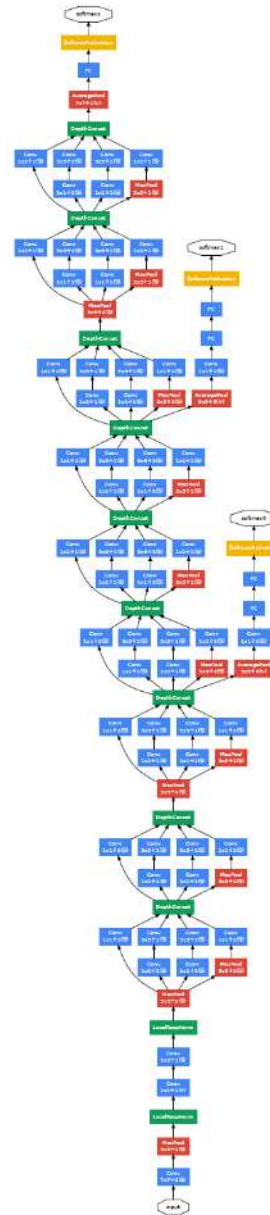
- Обучаем классификатор для 4000 человек – многослойную нейросеть
- 4М изображений – 4000 человек по 1000 изображений
- Последний слой используется как признаки для описания человека
- Точность 97.35% на базе Labeled faces in the wild (У человека на обрезанных – 97.53%)

Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf. **DeepFace: Closing the Gap to Human-Level Performance in Face Verification.** CVPR 2014.

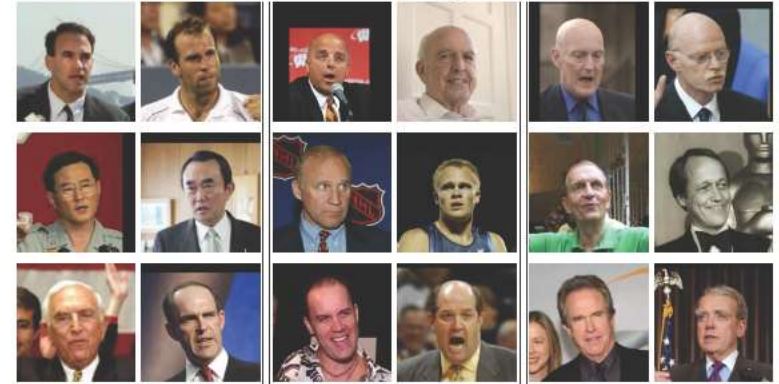


FaceNet от Google (2015)

- 220 млн. изображений
- 8 млн. разных людей
- От 6 до 140 млн. параметров в модели
- 99.63% точности на базе Labeled faces in the wild



False accept



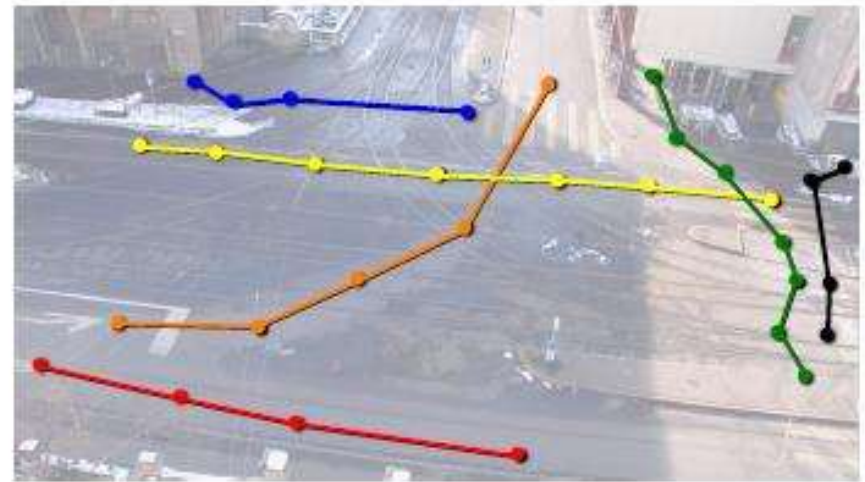
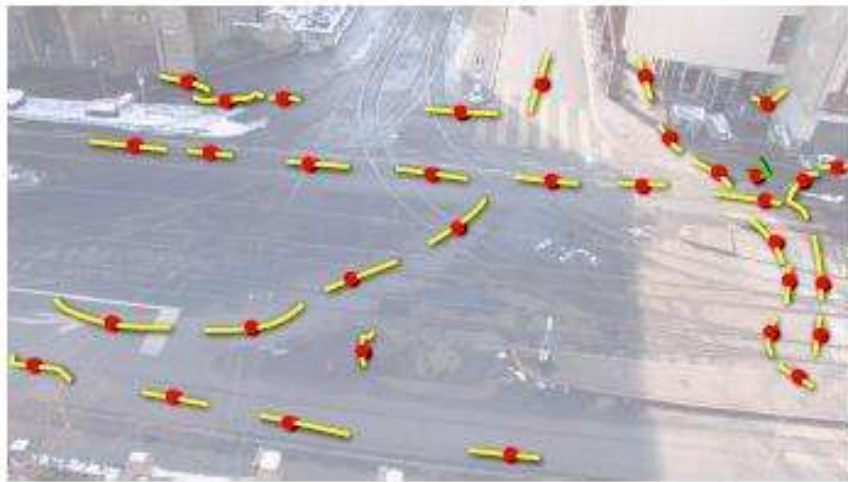
False reject





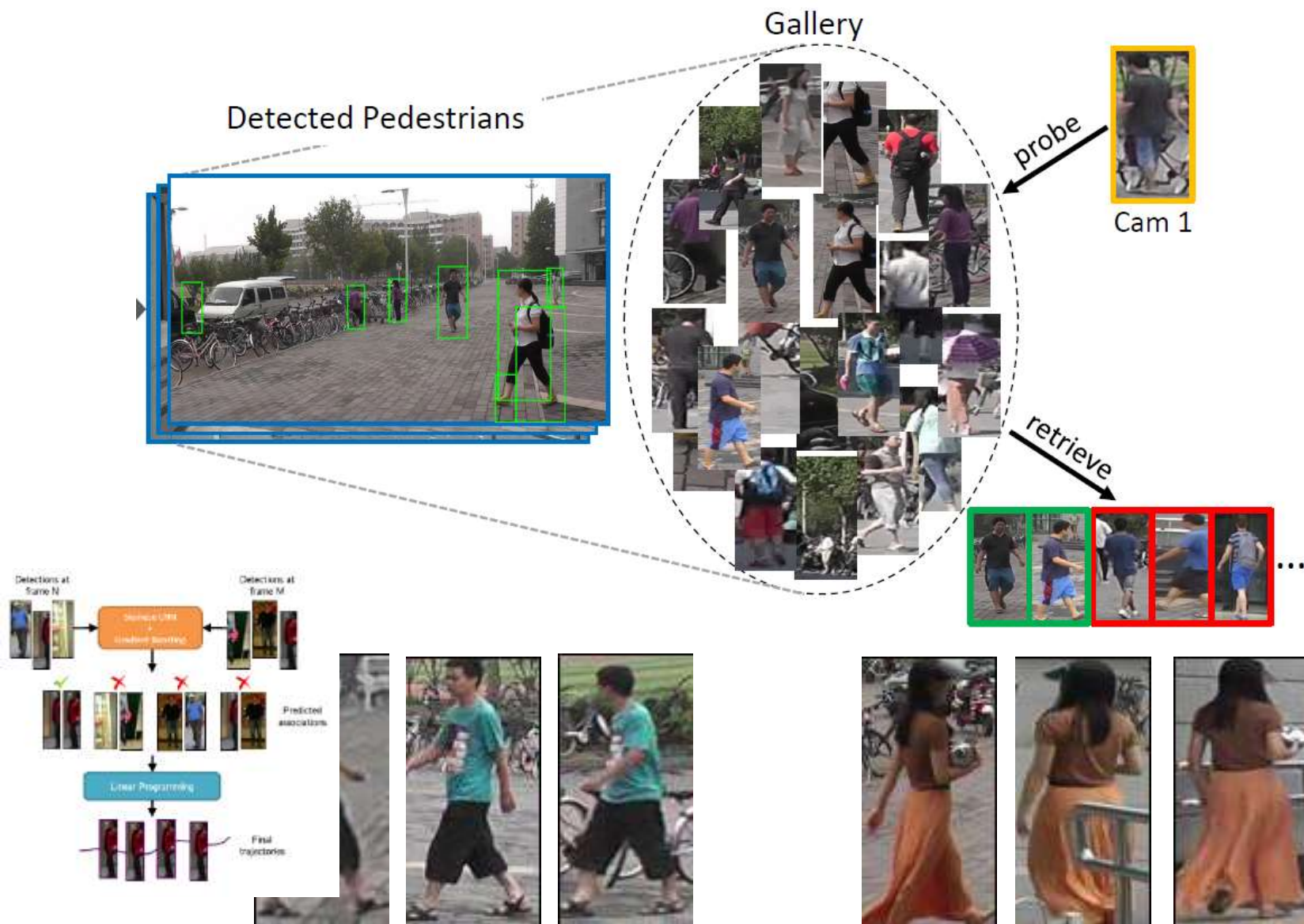
Восстановление из вектор-признака

Beyond: сопровождение объектов





Реидентификация людей





Поза человек, походка и события



Распознавание
позы нейросетью



Примеры из набора для
распознавания походки



Заключение

- Камер уже очень много, хотя покрытие пока далеко от 100%
- Камеры в основном записывают и хранят данные недолго (5 – 30 дней)
- Находить лица и распознавать научились весьма неплохо
 - Особенно крупные, поэтому – паспортный контроль, индексация в интернете и т.д.
 - В видео с камер видеонаблюдения – не так всё хорошо
- Пилотные проекты по умному наблюдению в разных странах и местах
- Очень много открытых задач



Кто этим занимается?



Наука и индустрия



«Учитель нейросетей»

- Программирование на Python, C/C++
- Базовая математика (линейная алгебра, математический анализ, численные методы, методы оптимизации, теория вероятностей и математическая статистика)
- Машинное обучение
- Нейросети (Глубинное обучение, Обучение с подкреплением)
- Приложения и развития
 - Компьютерное зрение
 - Анализ неструктурированных данных (при. текстов)
 - Обработка сигналов
- Методы и системы обработки больших данных