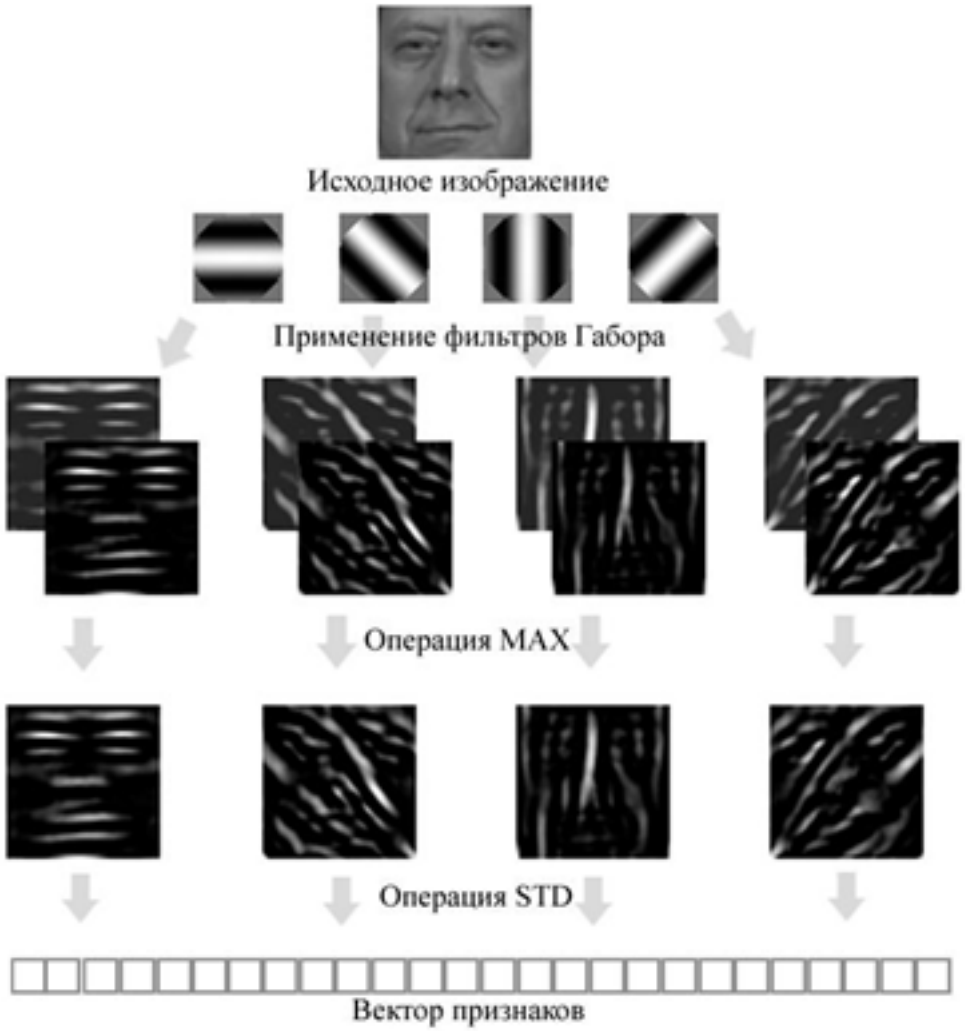
Результаты

- Обучали и тестировали на FERET
  - 10 сравнений – точность 80%
  - 20 сравнений – точность 90%
  - 1000 признаков – точность 96%

- Сравнение с SVM по скорости
  - В 25-50 раз быстрее

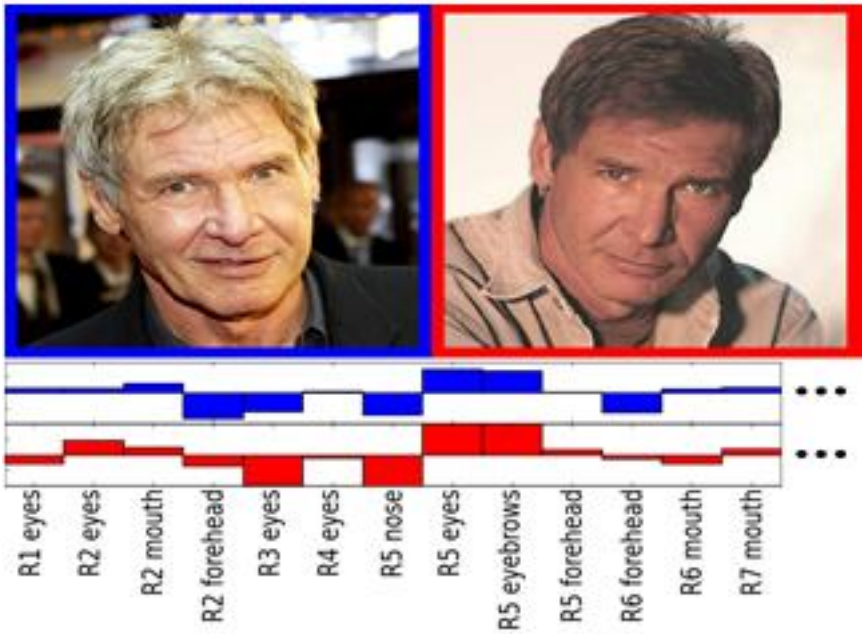
Bio-inspired features

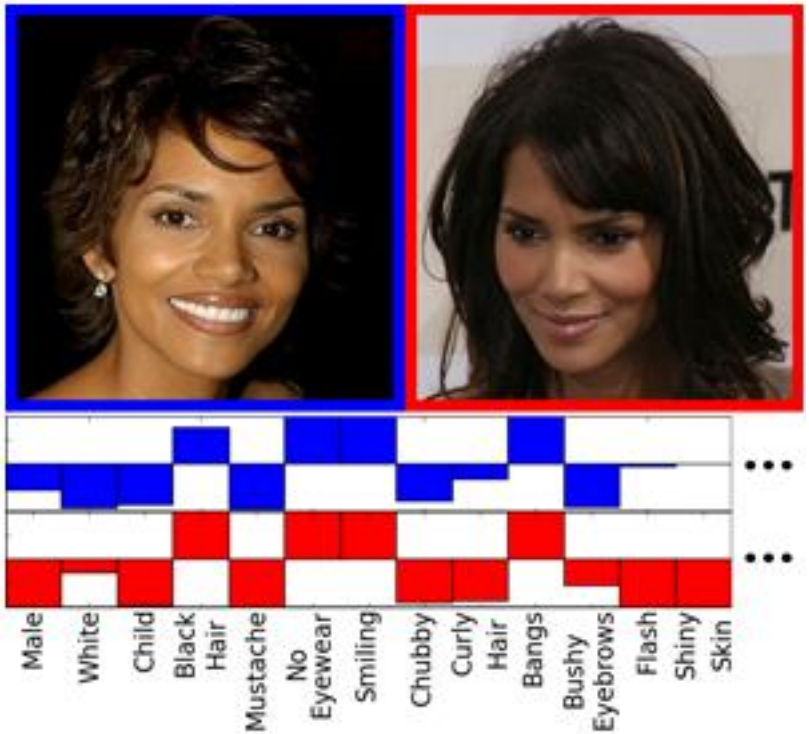


- Биологически-обусловленные признаки
- Фильтры Габора в 4х ориентациях и 16 масштабах
- Соседние масштабы объединяются через max-оператор
- Затем разбивается на пересекающиеся блоки и для каждого считается стандартное отклонение (STD)
- Всего ~3700 параметров

G-D. Guo, G. Mu, Y. Fu, and T. S. Huang, Human age estimation using bio-inspired features, CVPR 2009

Аттрибуты для верификации

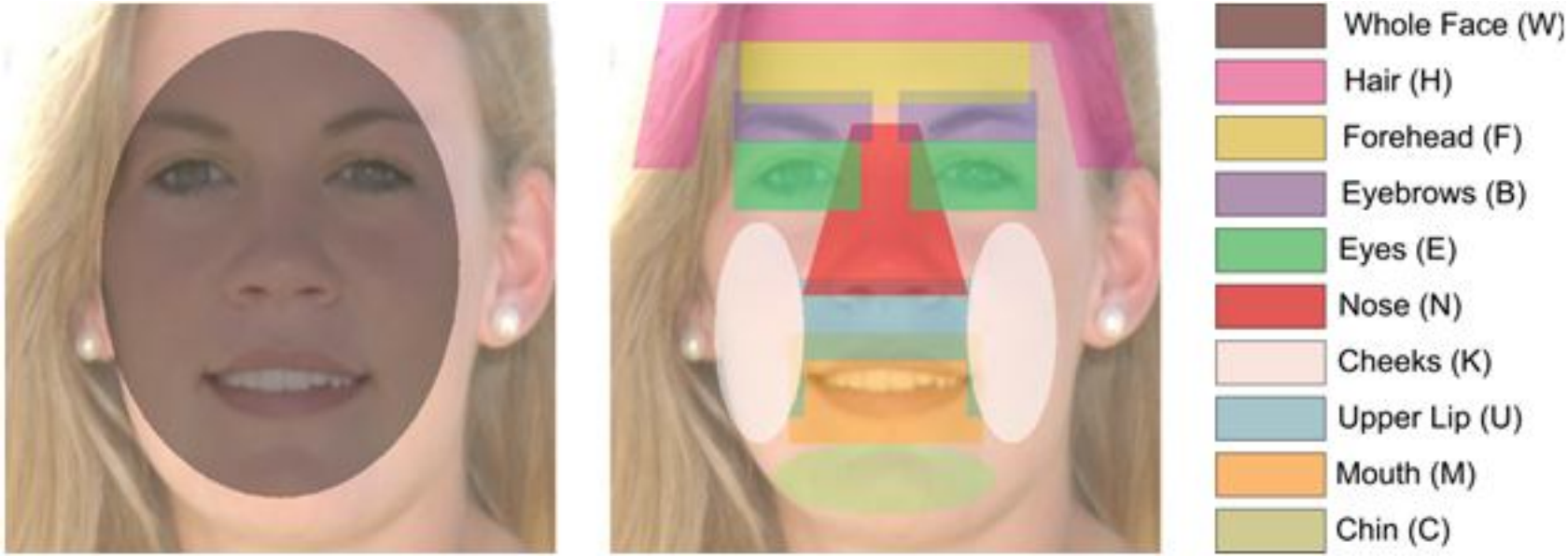




- Построим вектор-признак «высокого уровня»
- Набор атрибутов и «подобий»

Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, and Shree K. Nayar. **Attribute and Simile Classifiers for Face Verification.** *ICCV,* 2009.

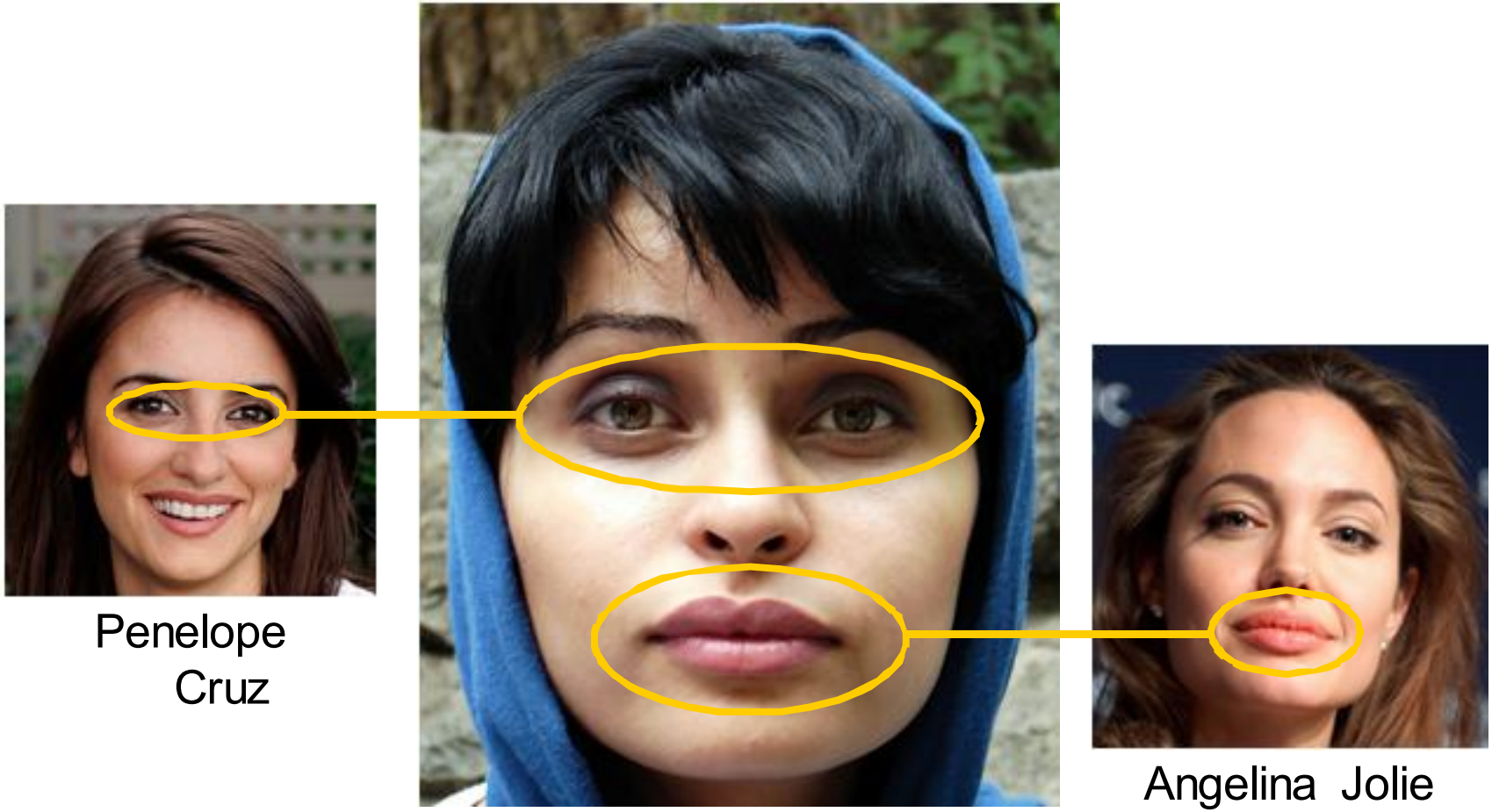
Классификаторы атрибута



Процедура построения классификаторов атрибута:

- Обучим классификаторы SVM на всех комбинациях (область, вид признак)
- Выбираем наилучшую комбинацию (область, признак)
- Выберем вторую наилучшую комбинацию, добавим в набор, обучим классификатор на наборе
- Будем добавлять новые наилучшие комбинации в набор, пока не прекратиться рост качества (или до 6 комбинаций)

«Similes» / «Подобия»



«Подобия» - классификатор, похожа ли данная часть лица на тестовом изображении и у другого конкретного человека

Обучение «подобий»

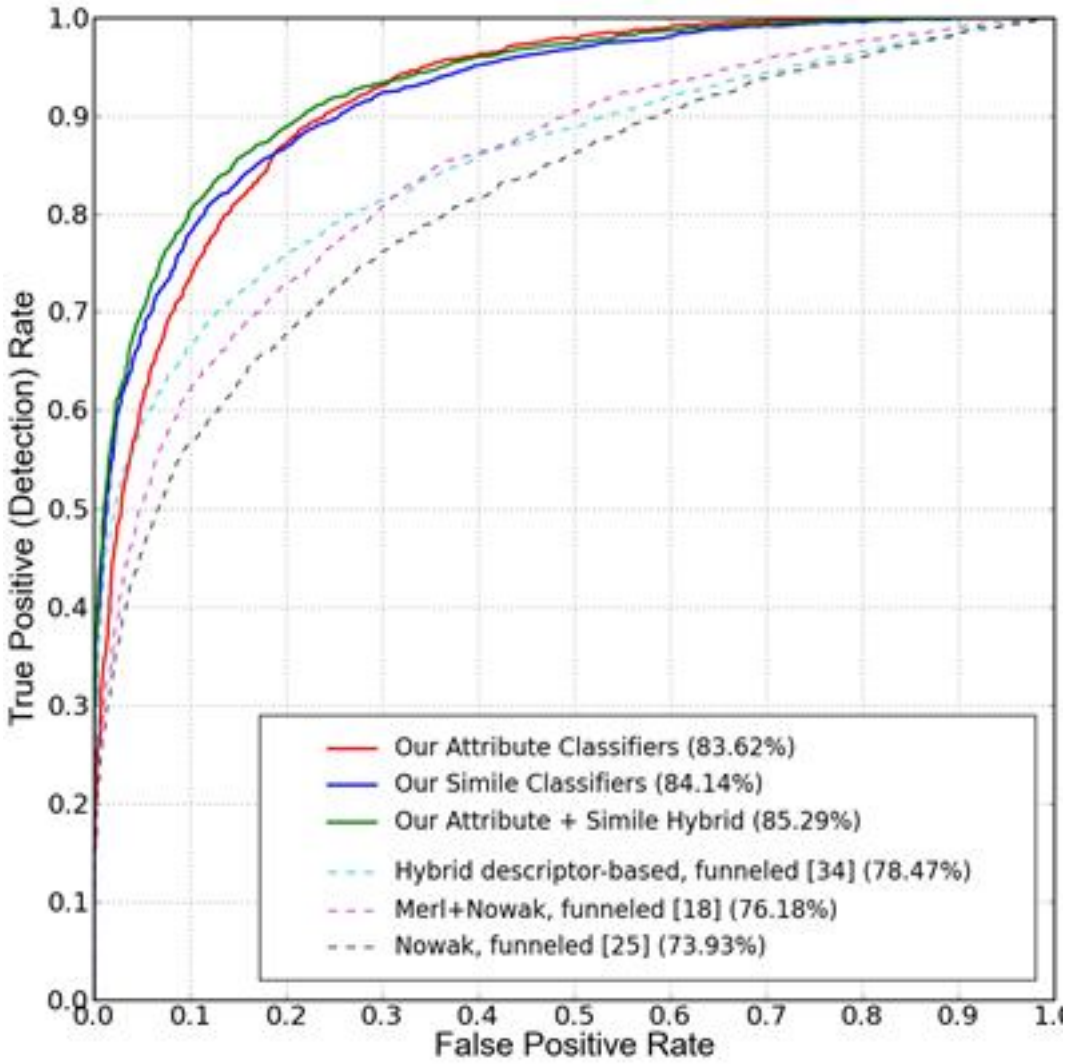
Изображения Пенелопы Круз (её глаз)

Изображения других людей (их глаз)

60 человек, 8 регионов, 6 типов особенностей (без выбора), для каждого обучаем классификатор

Алгоритм верификации

- Выходы всех классификаторов аттрибутов и similes объединяются в один вектор признаков.
- Этот вектор подается на вход новому SVM + RBF классификатору.



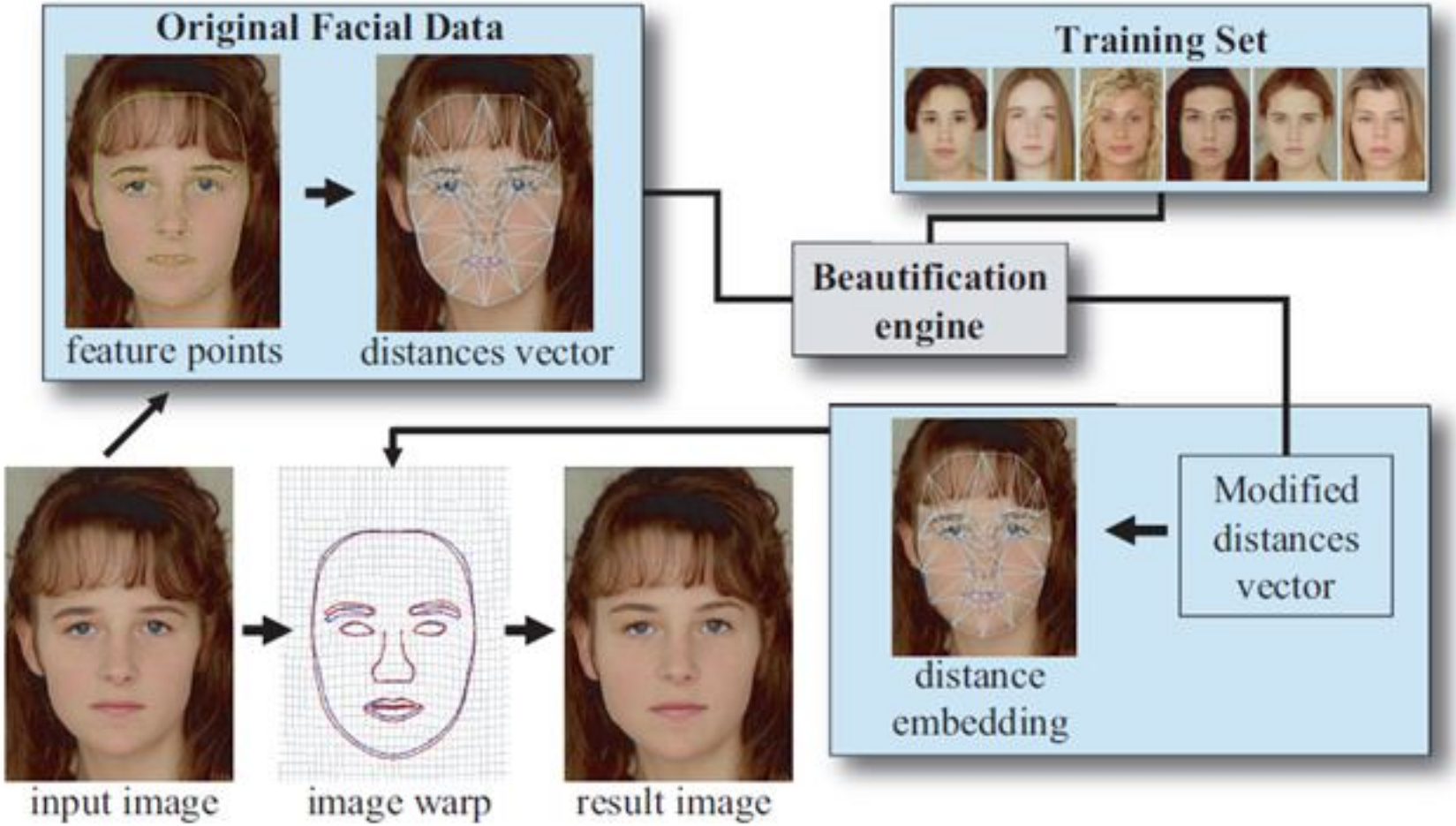
«Метод Тюбика»

- Эксперименты показали, что можно обучить классификатор SVM так, чтобы его результаты хорошо кореллировали с оценками людей (по шкале от 1 до 7 баллов)
- Попробуем чуть-чуть поправить все пропорции лица так, чтобы оно казалось красивее

EISENTHAL, Y., DROR, G., AND RUPPIN, E. 2006. Facial attractiveness: Beauty and the machine. Neural Computation 18, 1, 119–142.

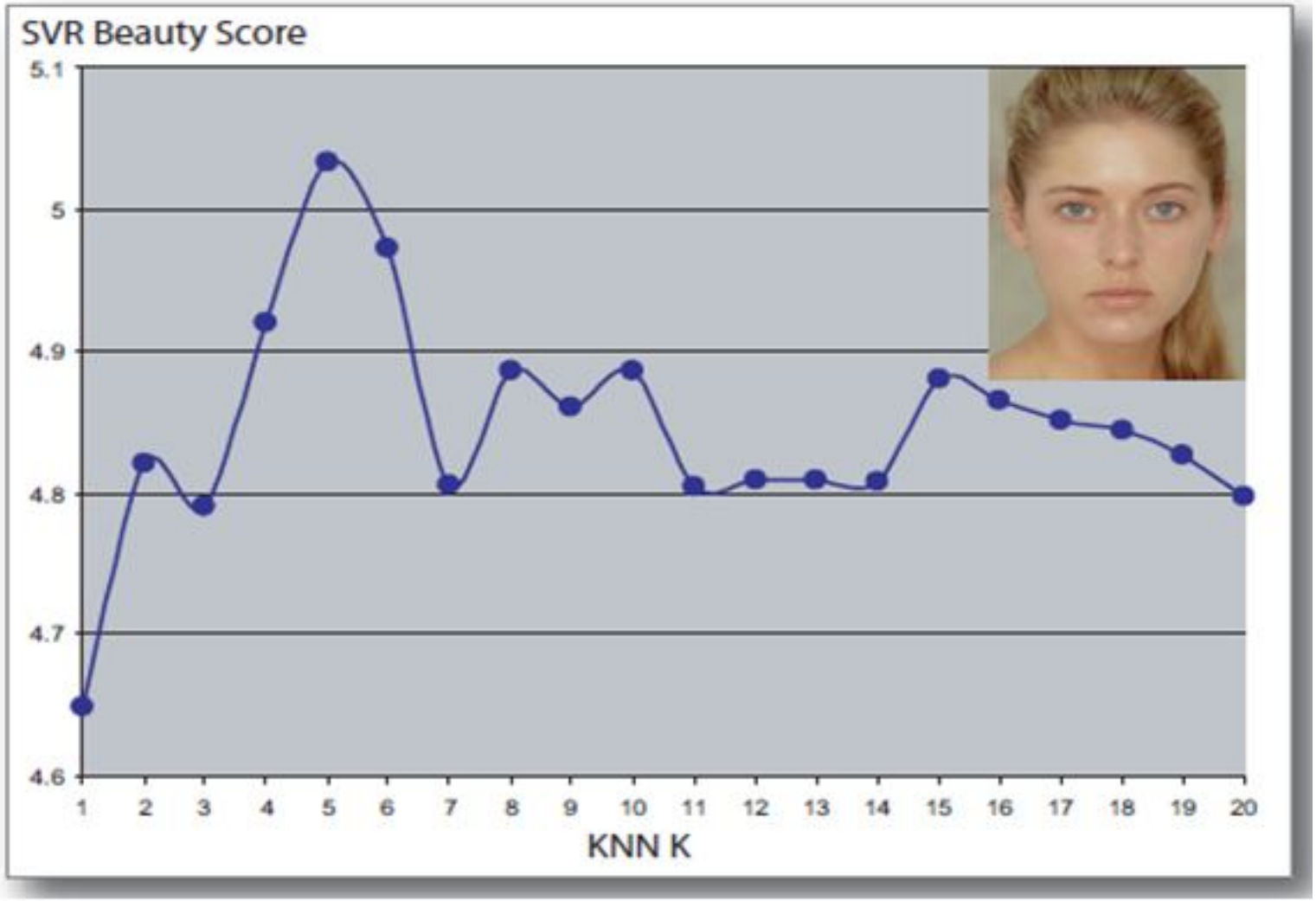
Tommer Leyvand, Daniel Cohen-Or, Gideon Dror and Dani Lischinski Data-Driven Enhancement of Facial Attractiveness ACM SIGGRAPH 2008

Схема



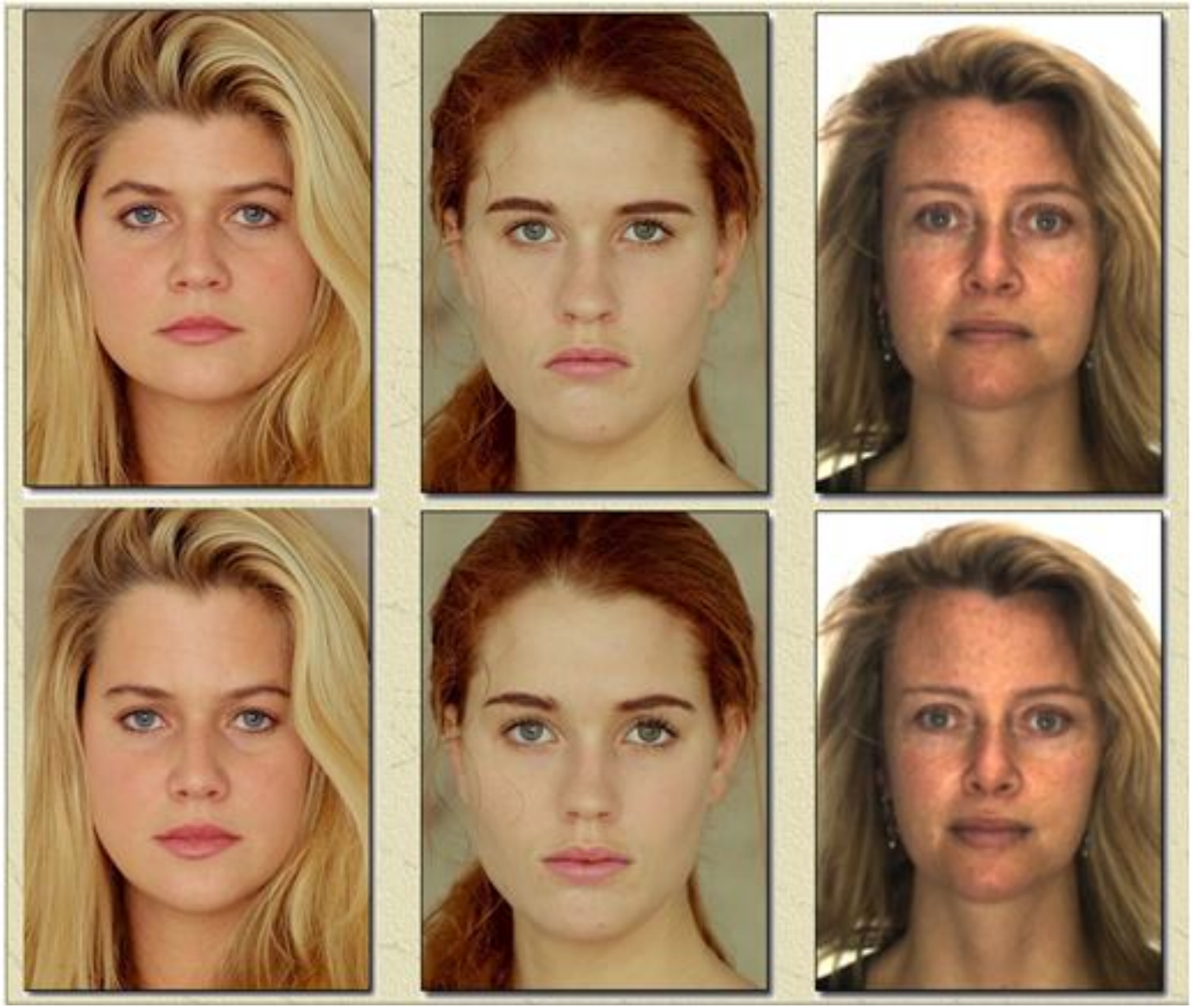
- Вычисление черт на основе ASM моделей

Зависимость от K в K-NN



- Один из методов – выбор нескольких похожих, но более красивых лиц и их комбинация

Результаты



Результаты

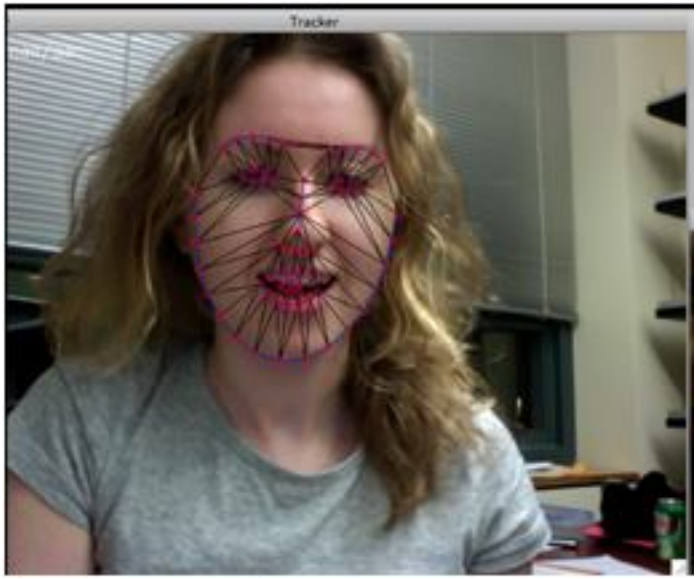
Результаты

Синтез анимации

- Идея – для текущего изображения лица найдём в базе изораженений / видео изображения с наиболее похожими выражениями

Ira Kemelmacher-Shlizerman, Aditya Sankar, Eli Shechtman, and Steven M. Seitz. Being John Malkovich, ECCV 2010

Выделение черт лица

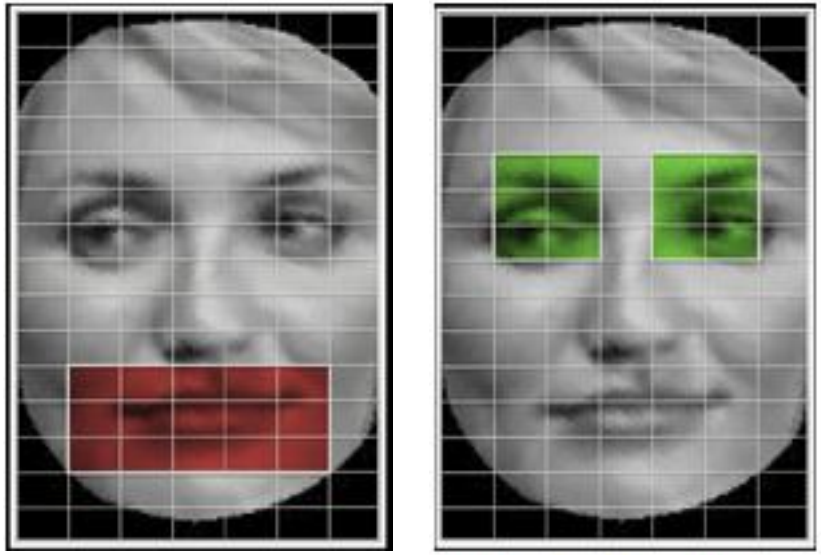


- Любые методы подойдут
- Морфируемые модели дадут наиболее точный результат

Канонические лица

- Лицо отображается на каноническую позу. Таким образом, мы убираем разницу в позе и можем сравнивать выражение лица напрямую

Расстояние между выражениями



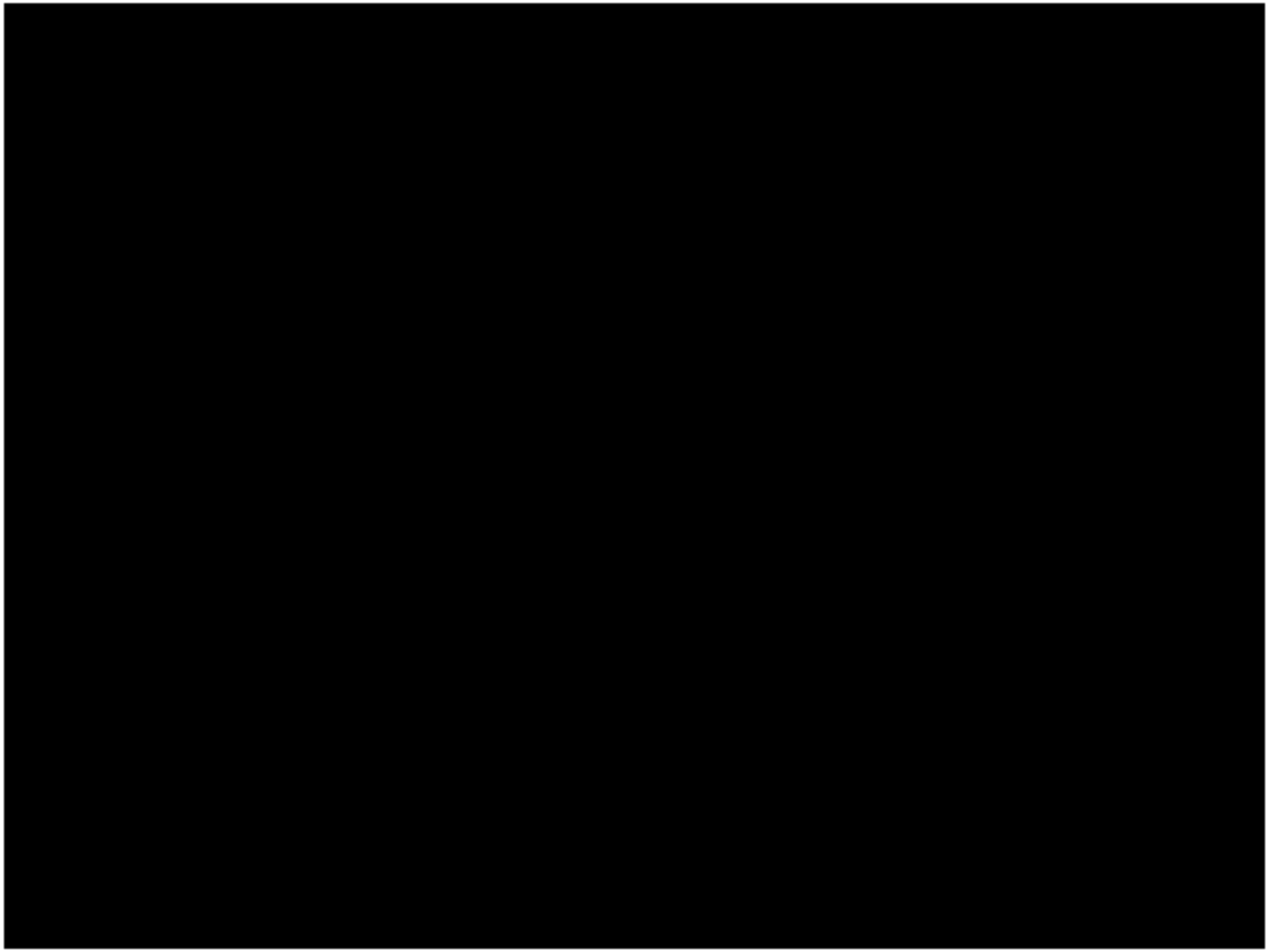
- Все изображение разбивается на блоки
- В каждом блоке считаем LBP
- Сравнивается область лица и область глаз

- Расстояние между блоками: $\chi^2(x,y) = 1/2 \sum_i (x_i - y_i)^2/(x_i + y_i)$

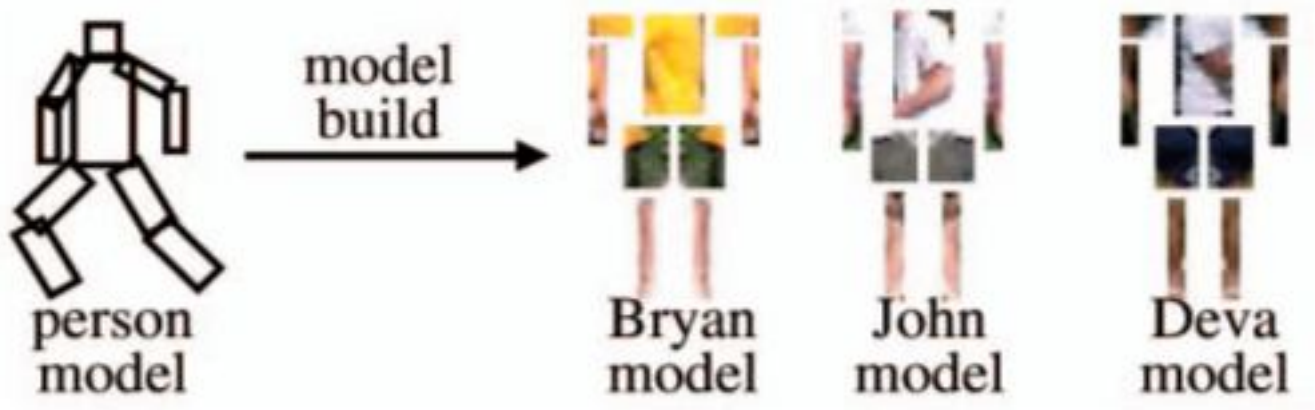
- Для каждой области суммируем расстояние между всеми блоками
- Суммарное расстояние в выражении лиц – взвешенная сумма расстояний по области рта и области глаз:

$$d_{appear}(i,j) \;=\; \alpha^m d^m(i,j) + \alpha^e d^e(i,j)$$

Видео



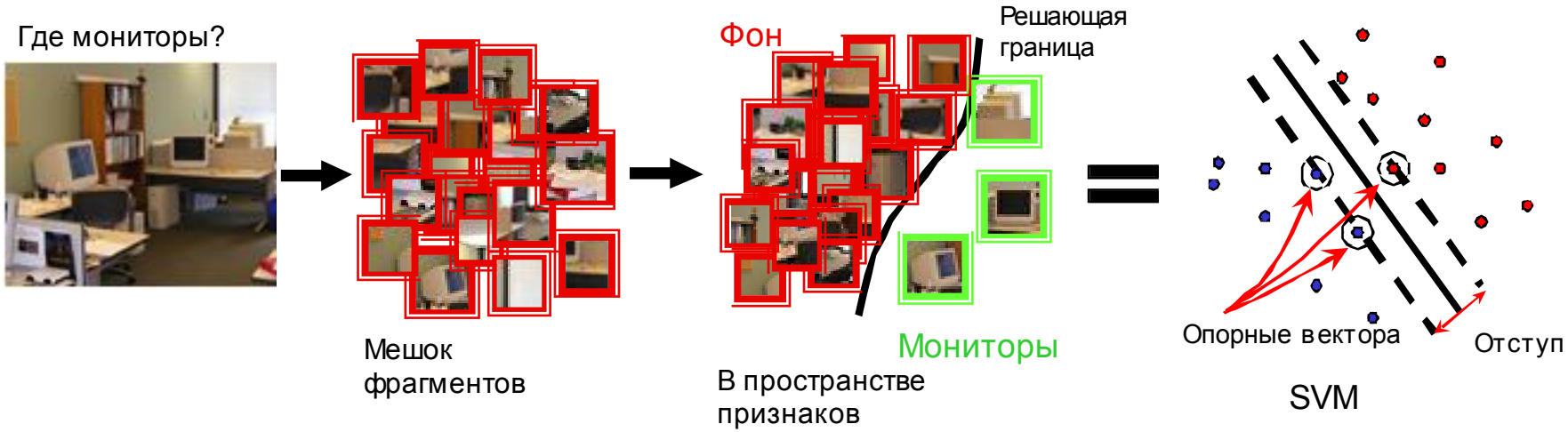
Модели из набора частей



Антон Конушин

Выделение пешеходов

Каким «стандартным» способом мы решаем эту задачу?



Что есть опорные вектора?

Выделение пешеходов

Что объединяет все эти изображения людей?

Похожие ракурсы, позы, и, как следствие, изображения

Вариабельность позы человека

Обратите внимание ещё на то, как похожи друг на друга по цвету/текстуре отдельные части тела человека

Сложность задачи

Сравним с поиском пешеходов:

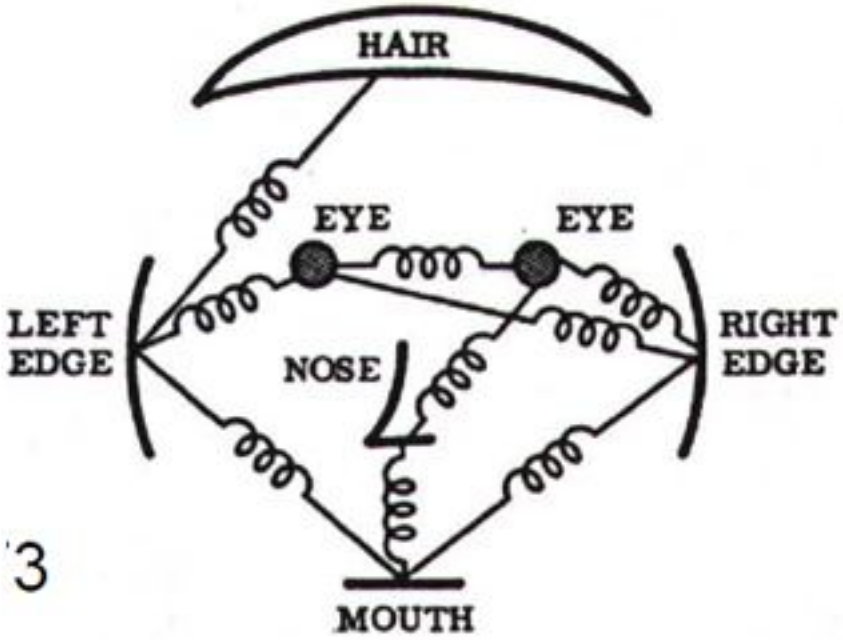
Похожие позы, ракурсы -> хорошо сработали «шаблоны всего человека»

Разные позы (сложные конфигурации), разные ракурсы -> шаблонов «человека» потребуется очень много для моделирования «в целом»

Модель объекта

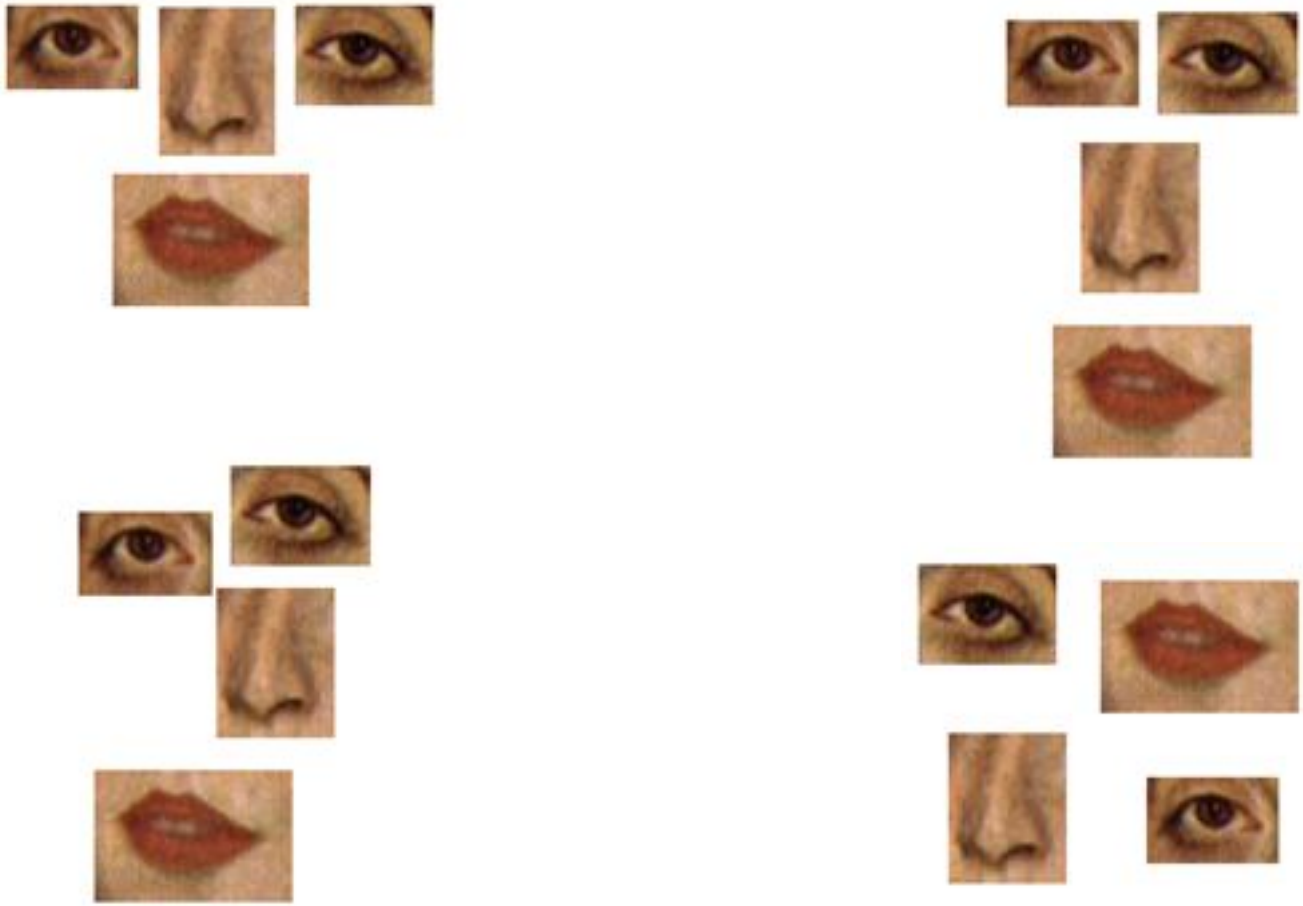
Иллюстрированные структуры (pictorial structures)

- Интуитивная модель объекта
- Состоит из 2х компонентов:
  - Частей (2D фрагменты изображения)
  - Структуры (конфигурация частей)
- Впервые была предложение Fischler & Elschlager 1973

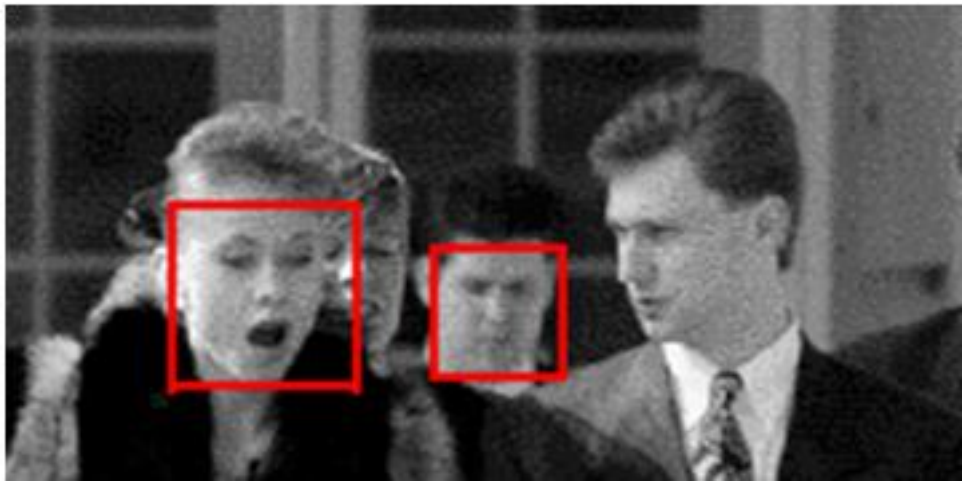


Ограничения на конфигурацию

Структура задает ограничение на взаимное расположение частей объекта



Причины деформации



Визуальные изменения из-за изменения ракурса

Внутренние деформации объектов

Учесть все деформации в одной модели может быть слишком тяжело! => может потребоваться несколько разных моделей

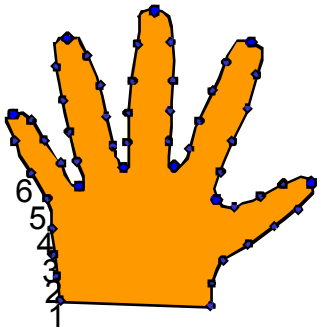
Части объекта

Антропометрические точки

Конечности человека

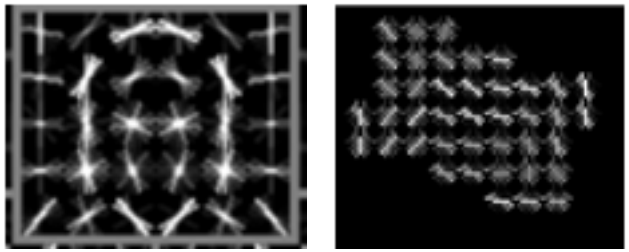
Можно как придумывать самим, так и доверить алгоритму автоматически выделить части

Детекторы частей

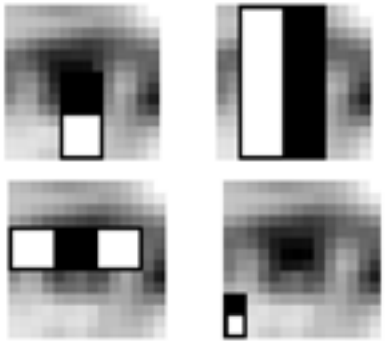


Точки контура

Локальные точечные особенности



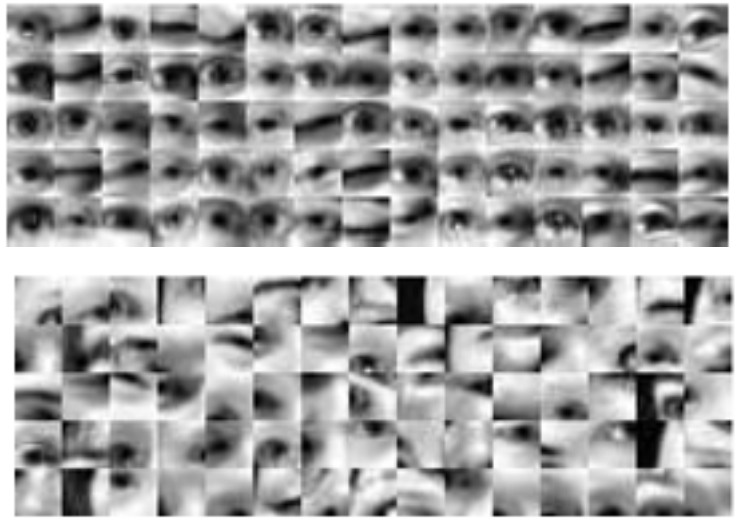
HOG-шаблоны



VJ-детектор

Опыт применения детекторов

Составные части объекта могут быть очень «неприметными»

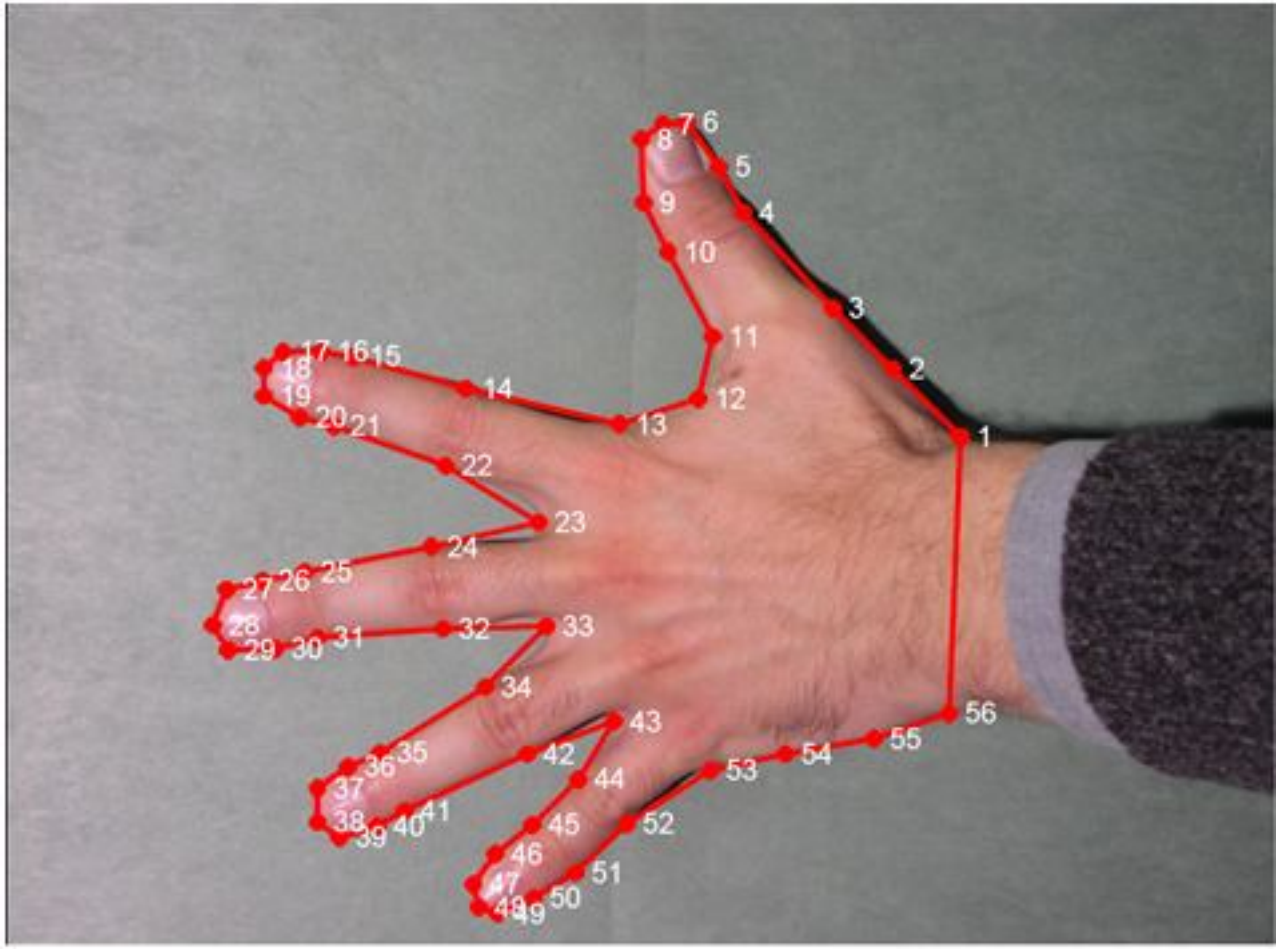


Положительные и отрицательные примеры для выделения глаза

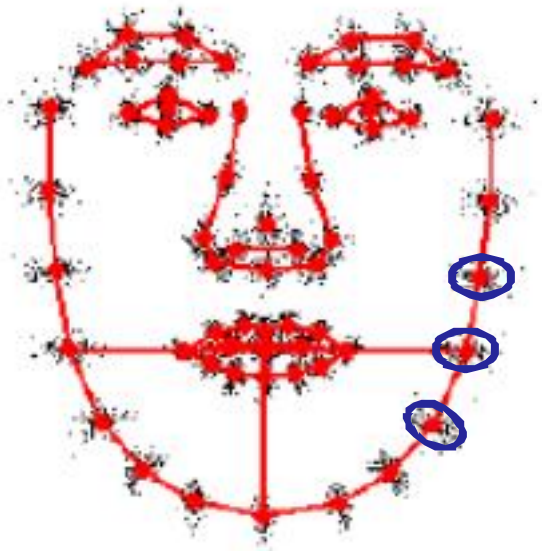
- Обычные алгоритмы выделения объектов (HOG + SVM, Haar + Boosting) не справляются с задачей локализации индивидуальных составных частей
- Локализовать «непримечательные» составные части мы можем только в совокупности, т.е. нужно моделировать зависимости

Построение модели

- Размечаем части нашего объекта
- В пример - контрольные точки на контуре ладони



Статистическая модель



- Идея: пусть все координаты всех точек составляют один вектор из некоторого многомерного распределения (2n, где n – число точек)
  - Пример - нормальное распределение
- Поскольку параметров слишком много, применим какой-нибудь метод уменьшения размерности
  - Метод главных компонент
- Получим уменьшенный вектор параметров

T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham, Training Models of Shape from Sets of Examples. // British Machine Vision Conference (BMVC). 1992

Пример: Лицо

Варьируем b_1

Варьируем b_2

Варьируем b_3

Пример: Рука

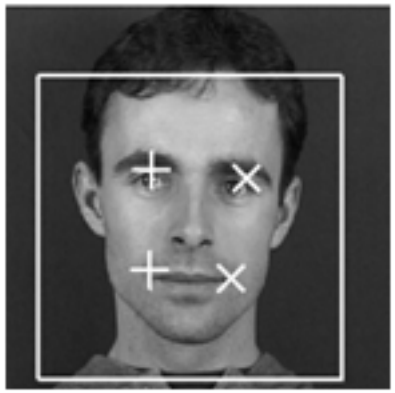
Варьируем b_1

Варьируем b_2

Варьируем b_3

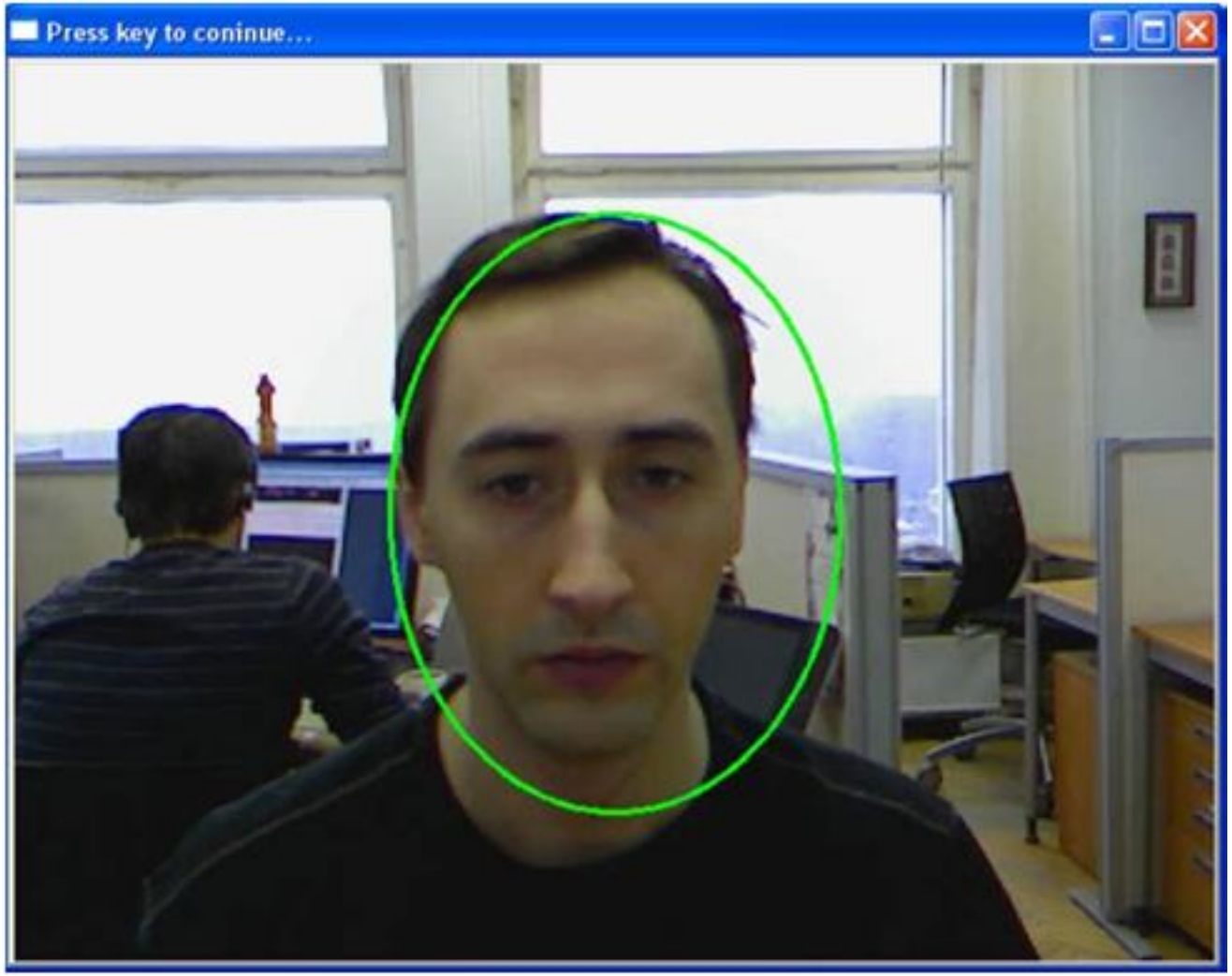
«Активная модель формы»

- Получили модель формы
- Как использовать ее для анализа изображений, т.е. для нахождения характерных точек?
- “Active Shape Model” - итеративный метод сопоставления модели изображению
  - Модель формы позволяет проверить возможность текущей конфигурации точек и исправить ошибки
  - Модель формы также позволит уточнить положение контрольных точек

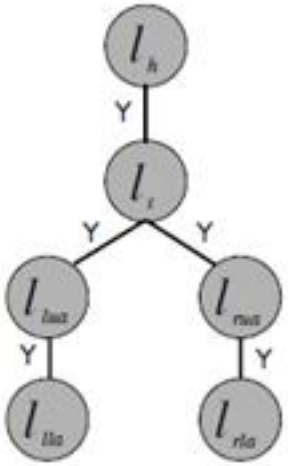


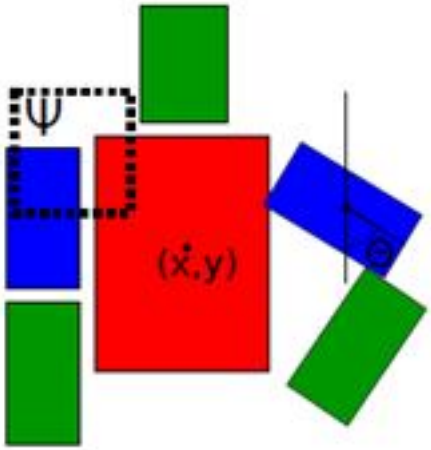
D. Cristinacce and T.F. Cootes, "Facial Feature Detection using ADABOOST with Shape Constraints", Proc.BMVC2003, Vol.1,pp.231-240

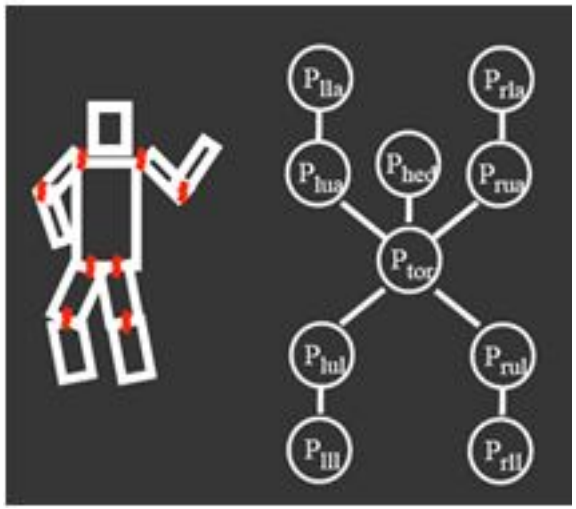
Пример работы алгоритма



Иллюстрированная структура





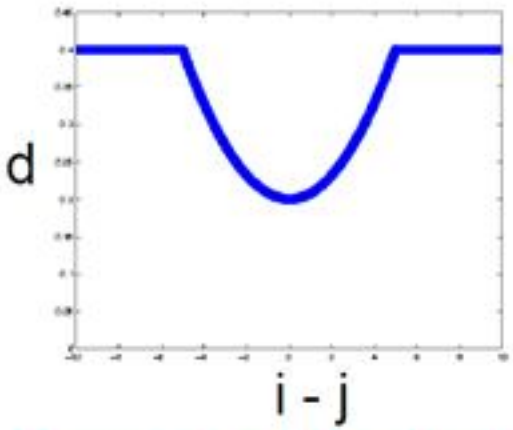


- Графическая модель по положениям частей на изображении

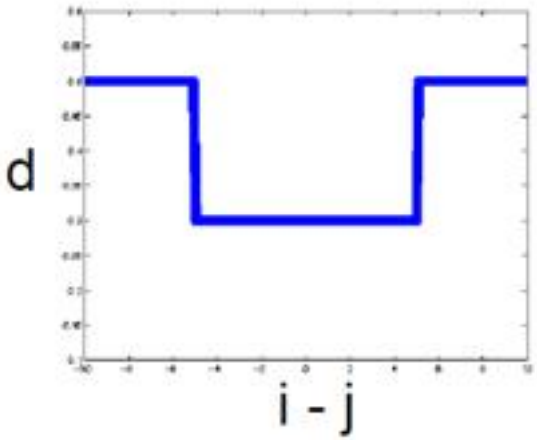
$$P(L|I,\Theta) \propto \exp\left(\sum_{(i,j)\in E} \Psi(l_i,l_j) + \sum_i \Phi(I|l_i,\Theta)\right)$$

- Φ – модель внешности части (на основе каких угодно признаков)
- Ψ – кинематическое ограничение

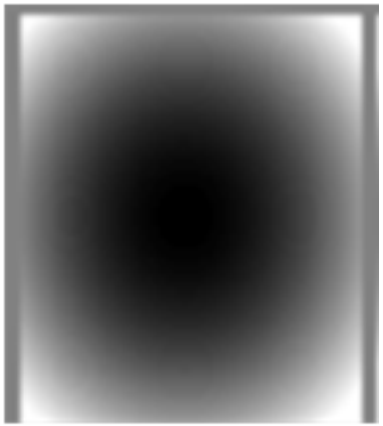
Кинематические ограничения



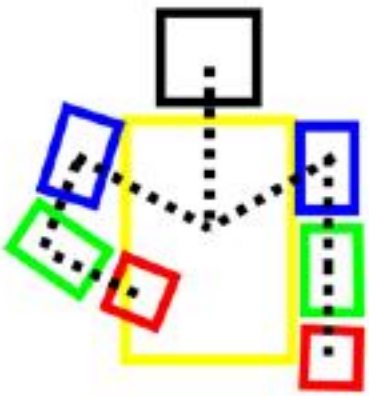
Truncated Quadratic



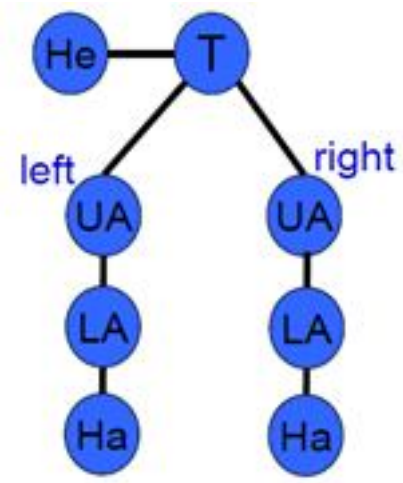
Potts



$$\theta_{ab;ij} = w_{ab} d(|i-j|)$$



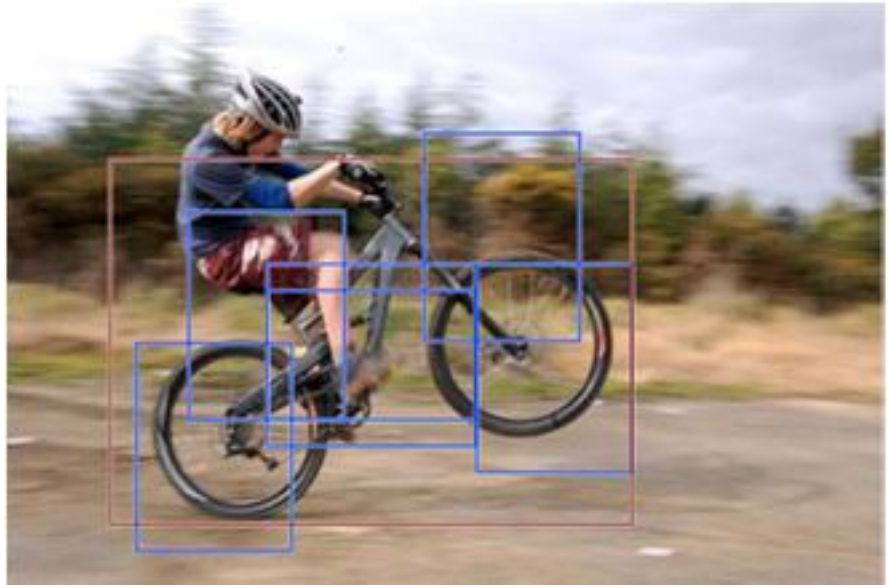
Вывод в модели

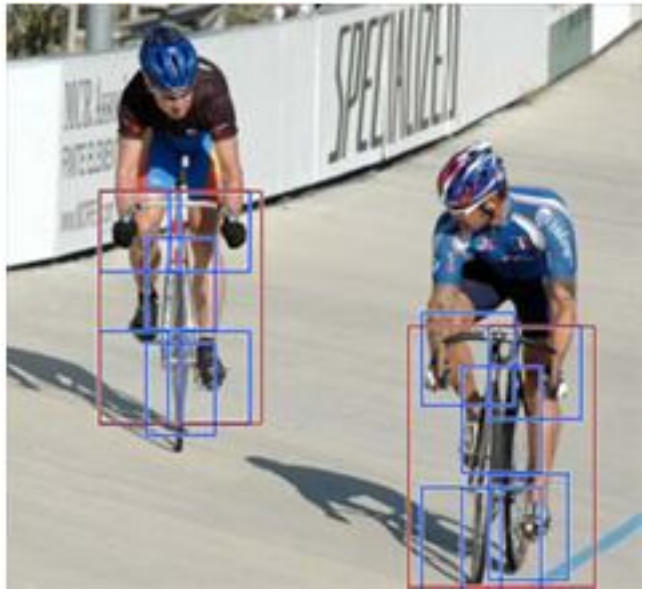


- Пусть h – количество вариантов положения каждой из n частей модели (после дискретизации)
- Тогда всего вариантов h^n
- Благодаря дереву, сложность оценки с помощью динамического программирования снижается до $O(nh^2)$
- Сложность достигает $O(nh)$ при использовании специальных ускорений

P. Felzenszwalb, D. Huttenlocher **Efficient Matching of Pictorial Structures** // IEEE CVPR, 2000

Известный детектор

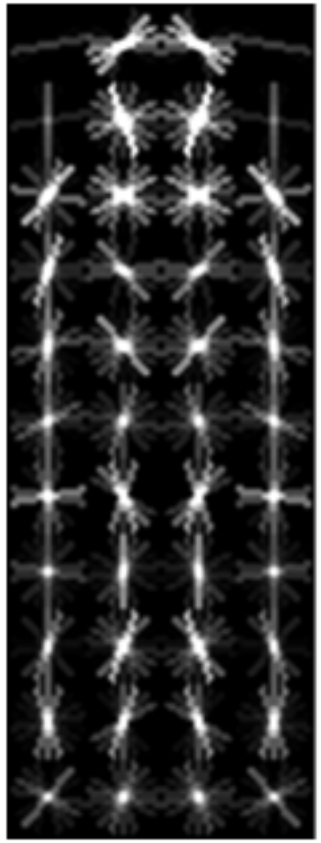




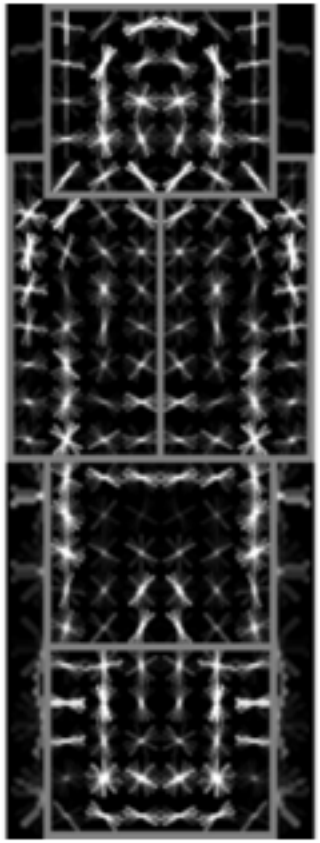
- Несколько моделей для разных ракурсов
- Каждая модель состоит из общего шаблона (HOG) и набора частей
- Для обучения достаточно рамок (bbox) объектов

Pedro F. Felzenszwalb et al. Object Detection with Discriminatively Trained Part Based Models. PAMI 2010

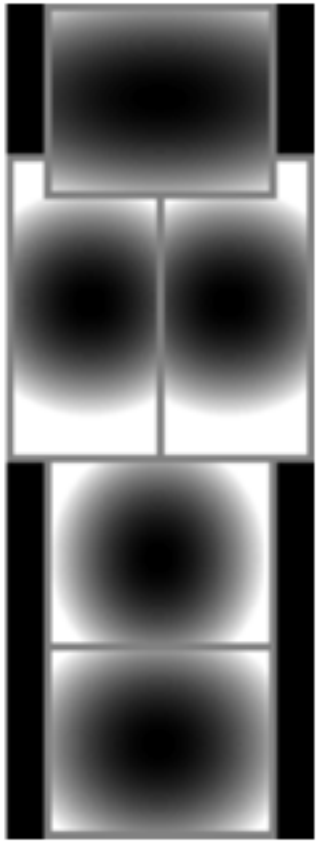
Модель



Корневой фильтр
Низкое разрешение



Фильтр части
2x разрешение



Модель
деформации

Функция качества

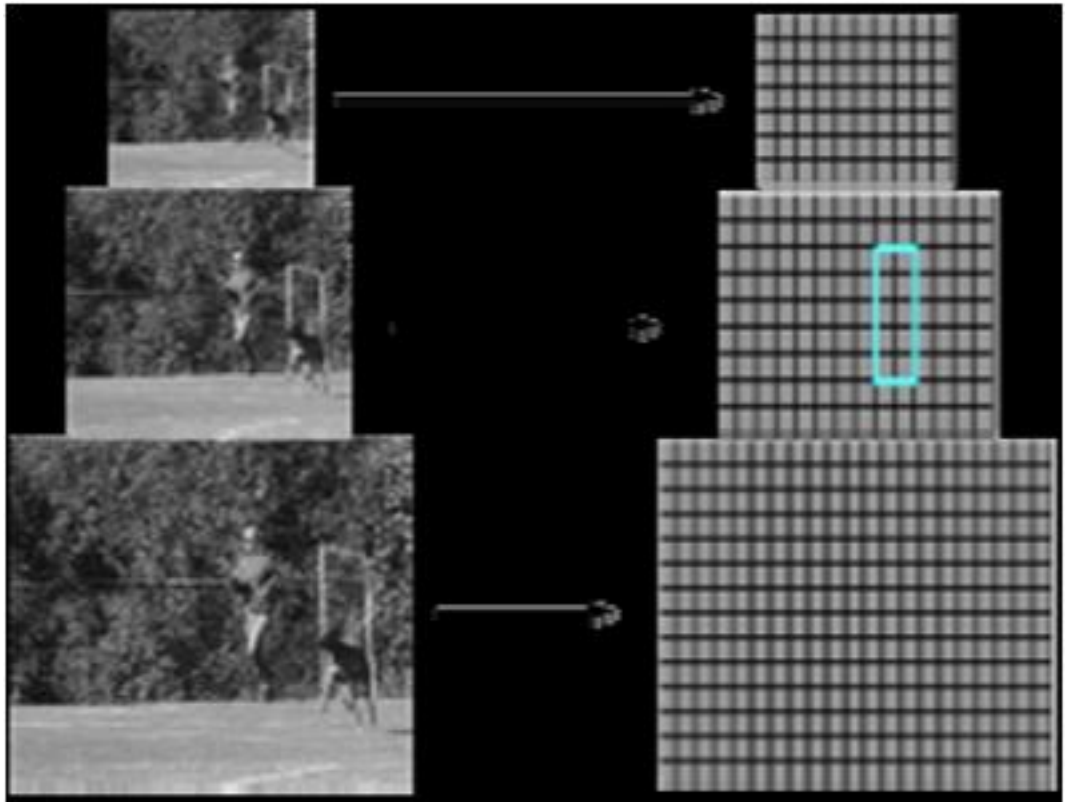
$score(x,z)$	$= \sum_i w_i \, \phi(x, z_i) +$	$\sum_{i,j} w_{ij} \, \Psi(z_i, z_j)$
x = image $z_i = (x_i, y_i)$ $z = \{z_1, z_2 \ldots\}$	Качество (Score) за локальные шаблоны частей	Модель деформации «пружинки»

Функция качества линейна по локальным шаблонам w_i и параметрам пружинок w_{ij}

$$score(x,z) = w \cdot \Phi(x, z)$$

Source: Deva Ramanan

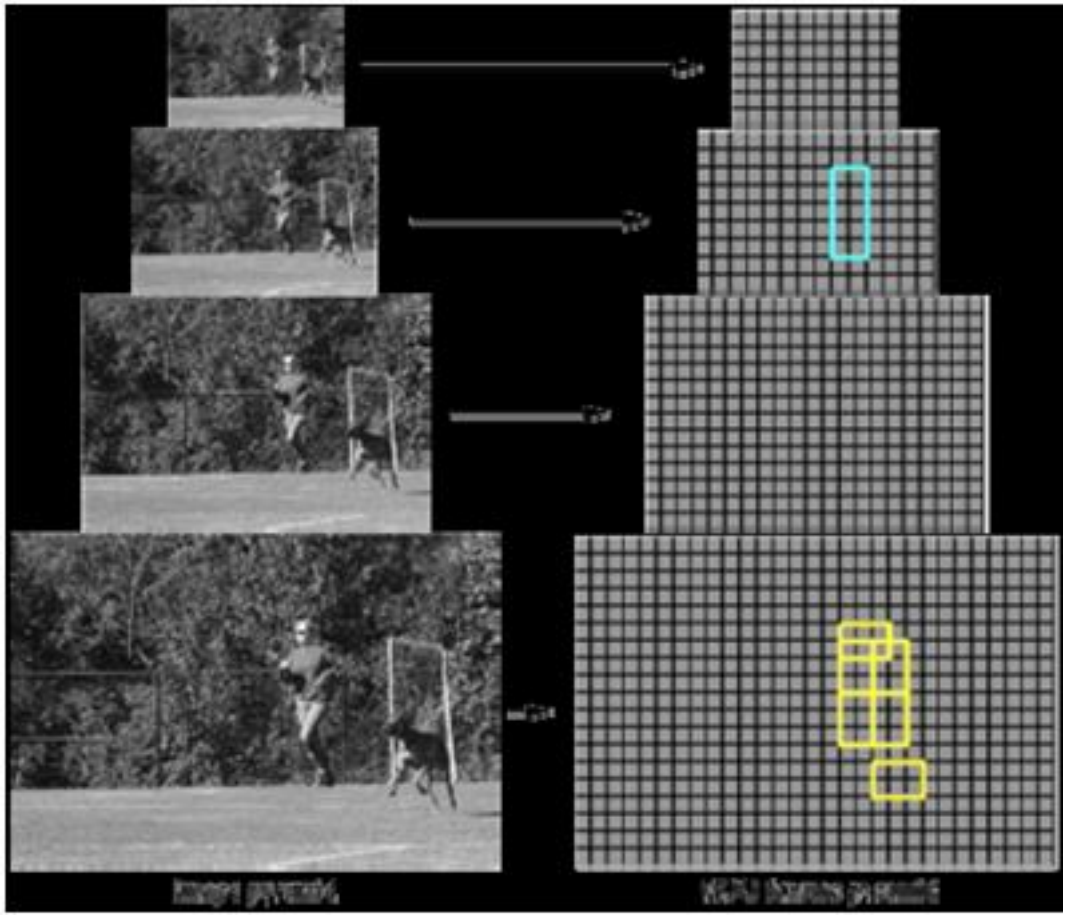
Корневой детектор

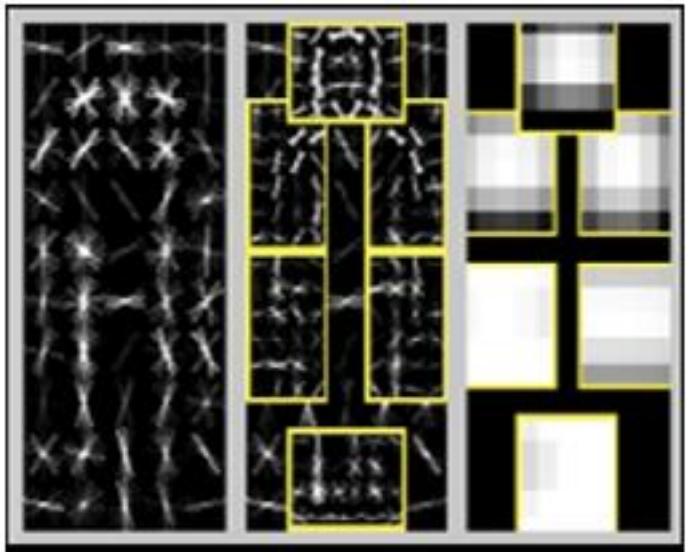


Корневой фильтр

- Скользящее окно по положению и масштабу
- Детектор части – HOG + линейный SVM
- Выбираем точки, где корневой детектор дал заметный отклик

Гипотезы частей



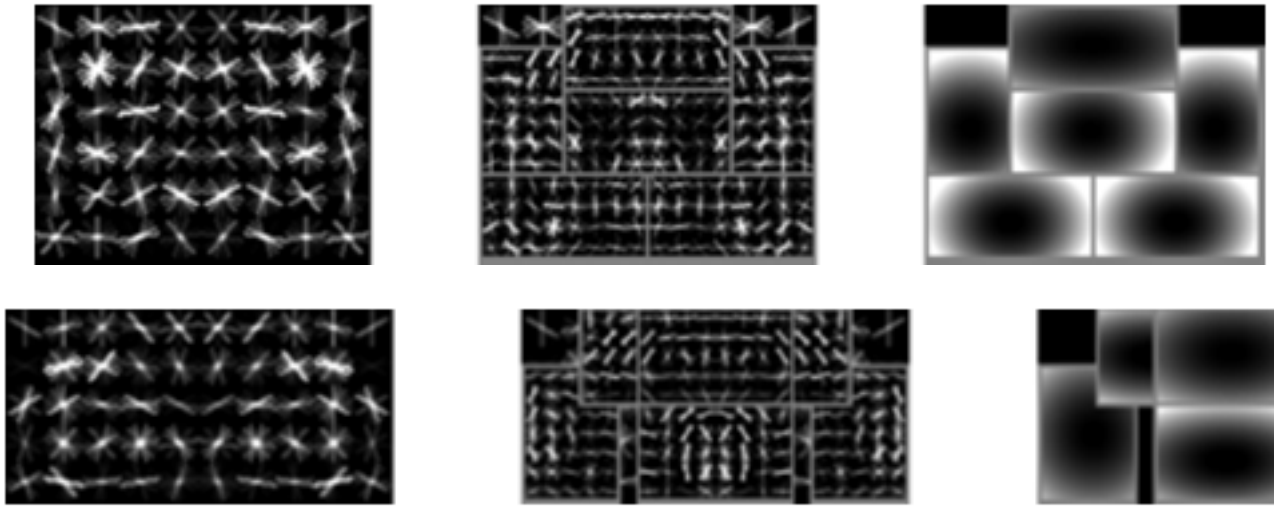


«Качество» гипотезы это сумма качеств частей минус пенальти за деформацию

- Начинаем искать части на 2x разрешении корневого фильтра
- Выбираем оптимальную конфигурацию

Модели для разных объектов

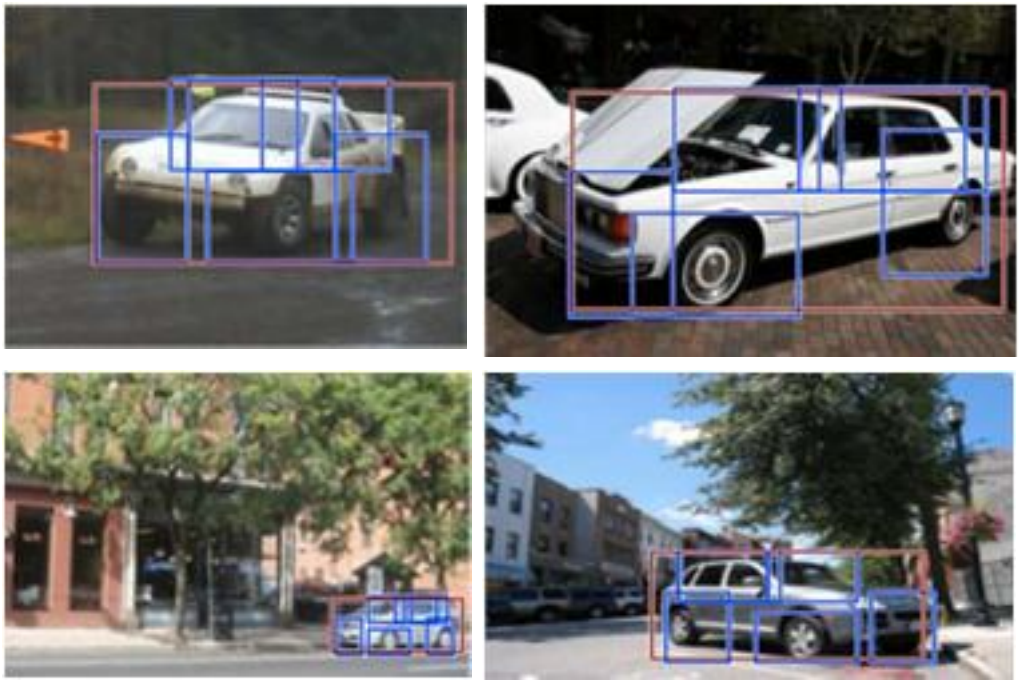
Собака



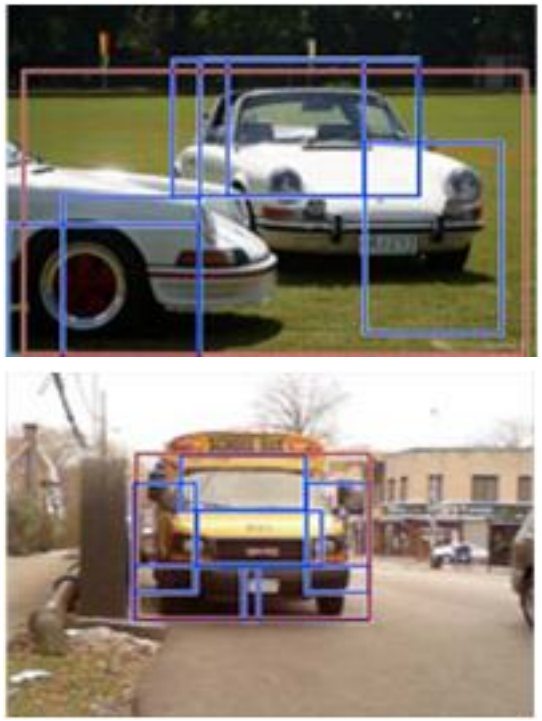
Автомобиль

Результаты на машинах

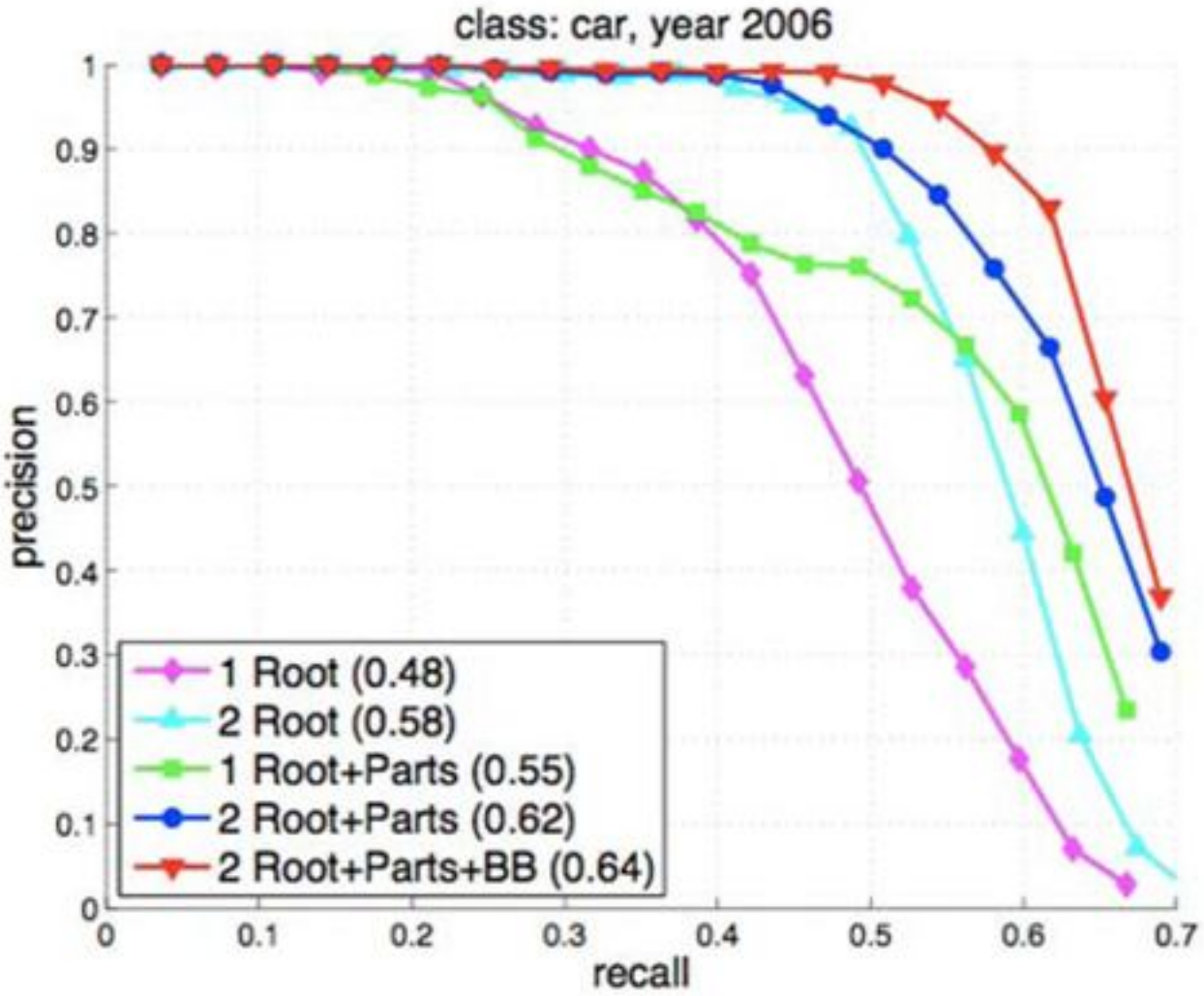
Правильные обнаружения с высокой оценкой



Ложные обнаружения с высокой оценкой

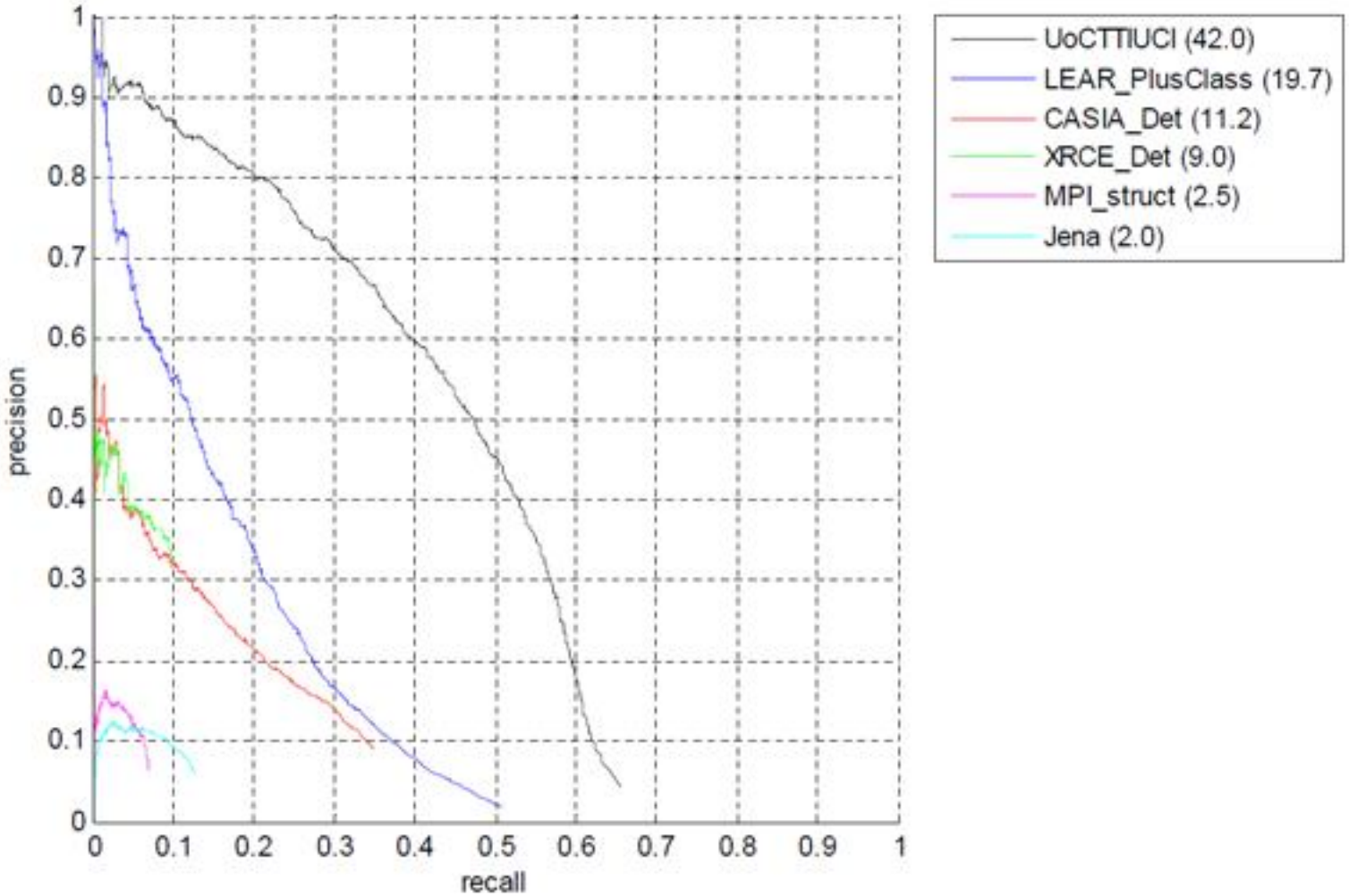


Сравнение моделей



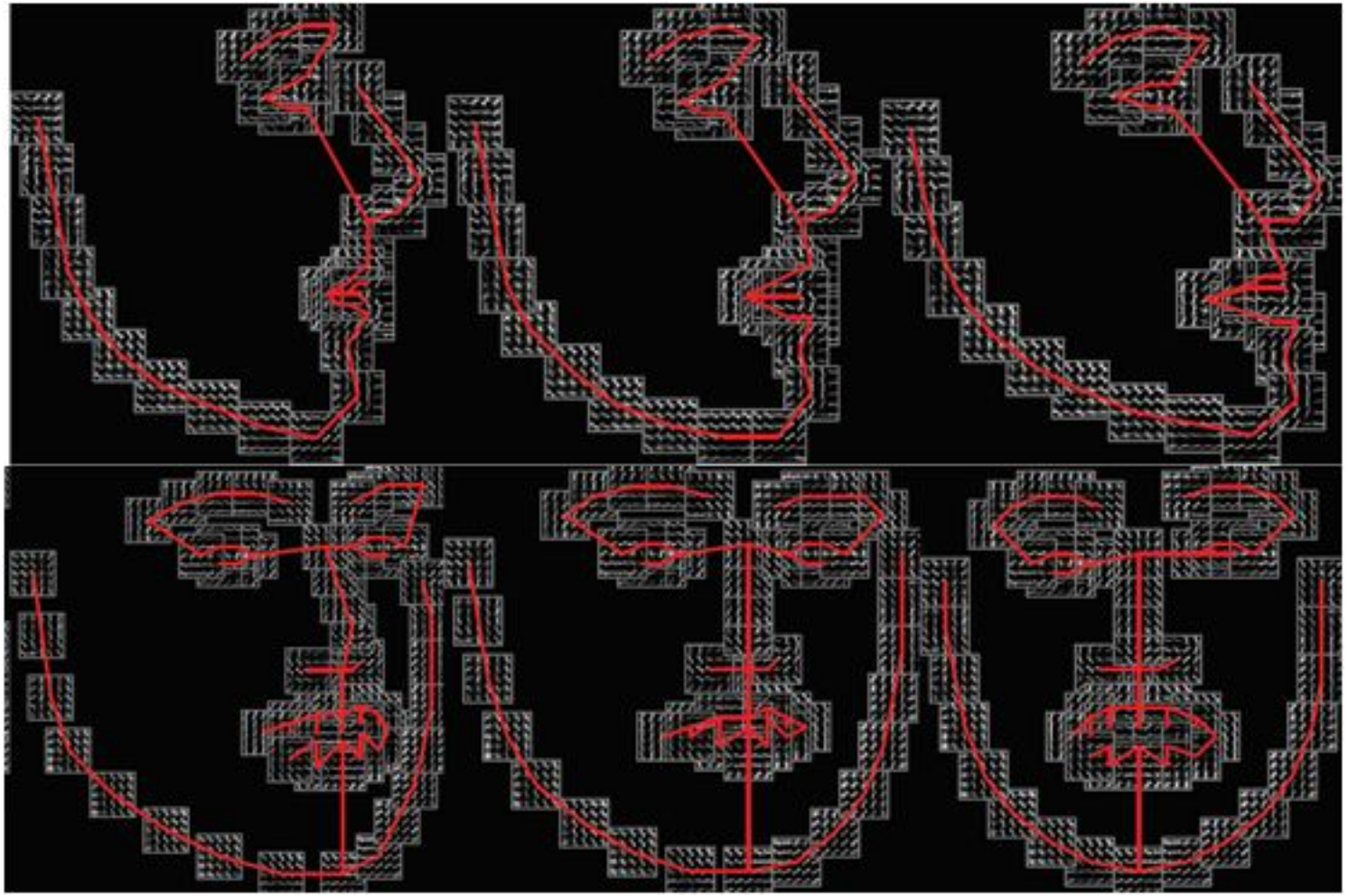
Модели для разных ракурсов и части – всё даёт свой вклад

Результаты



Эффект на момент появление метода

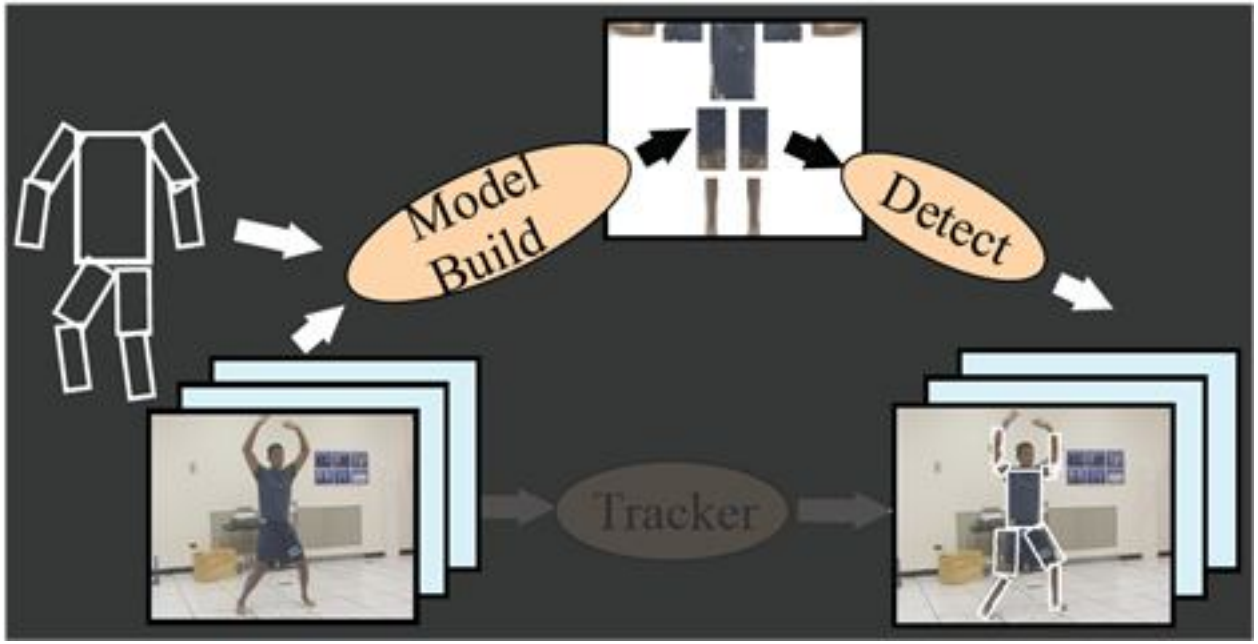
Пример для лица человека

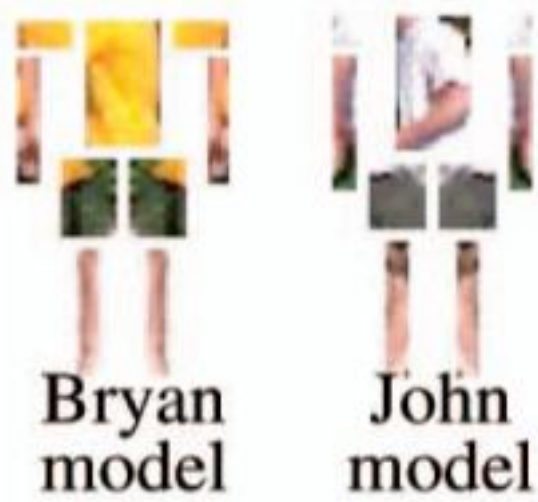


X. Zhu, D. Ramanan. "Face Detection, Pose Estimation, and Landmark Localization in the Wild" *Computer Vision and Pattern Recognition* (CVPR), Providence, RI, June 2012.

Результаты

Усложнение модели

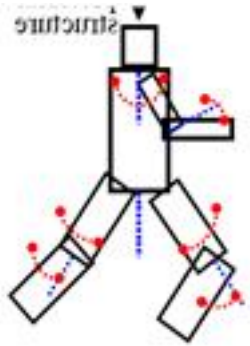




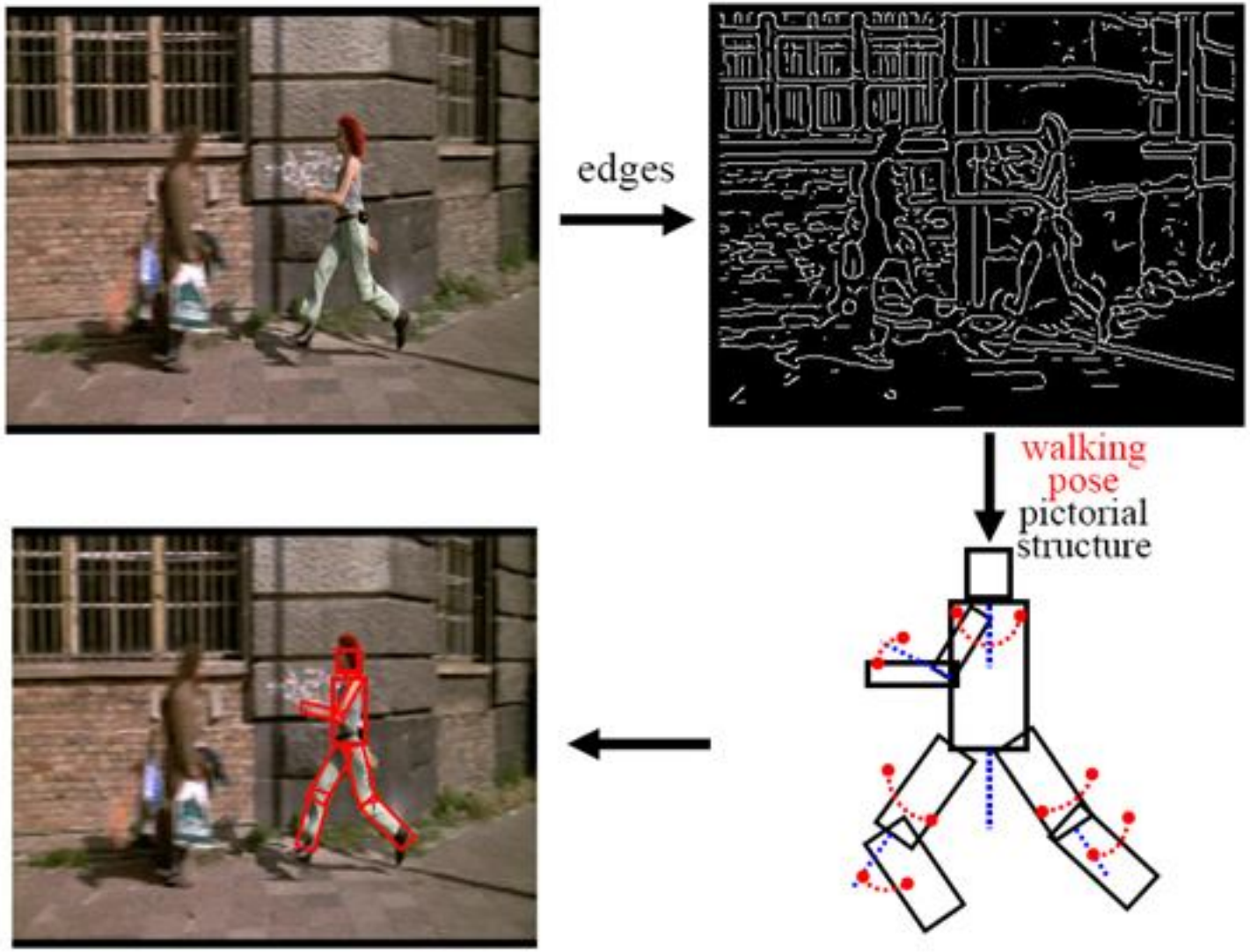
- Такие признаки, как цвет одежды, могут очень помочь при распознавании человека
- Но цвет одежды человека заранее неизвестен
- Выход – попробовать итеративно уточнять модель
  - Чем-то концептуально похоже на GrabCut

«Opportunistic detector»

- Будем искать «удобные для обнаружения» модели
- «Идущий человек» - хорошо
- «Падающий в прыжке человек» - плохо

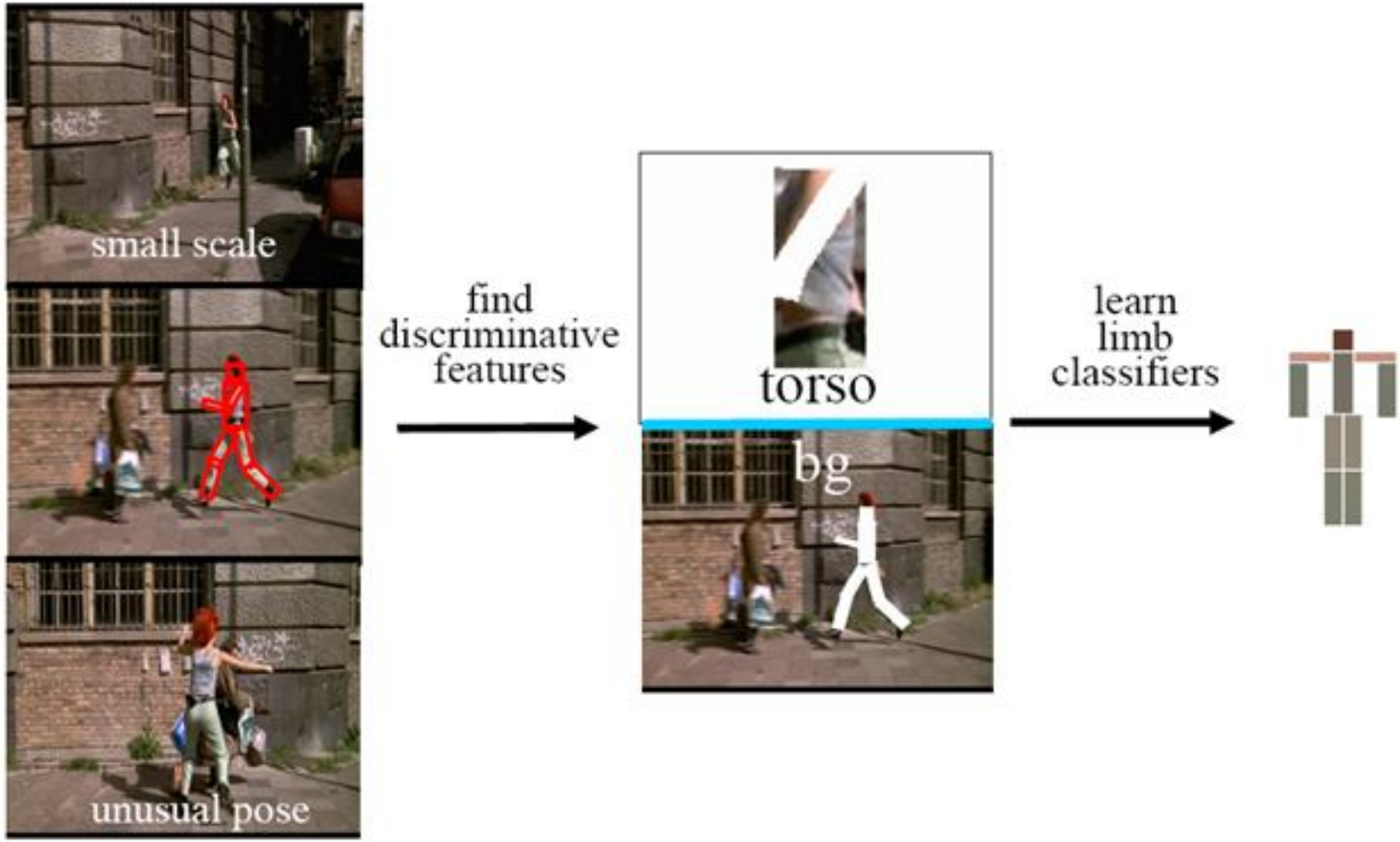


Детектор шагающего человека



Независимо считаем для краев разной ориентации chamfer distance, затем суммируем

Построение модели внешности



- Обучаем классификатор внешности (в RGB пространстве)
- quadratic logistic regression

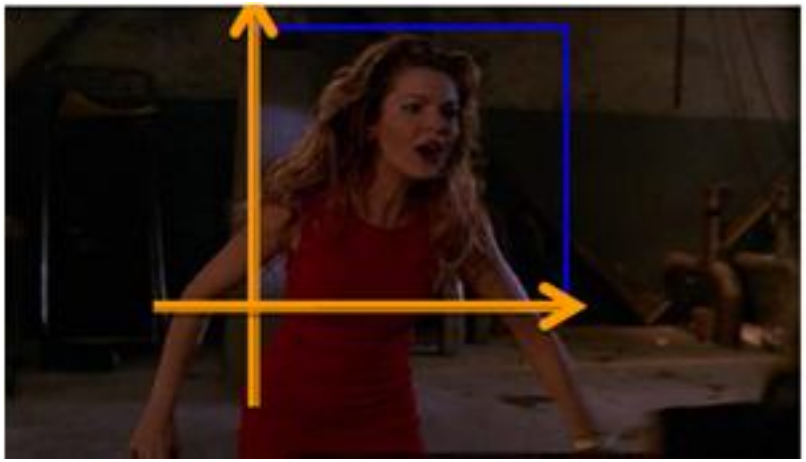
Пример работы

Обычный детектор

- Пример выборки отдельных кадров из сериала Buffy - разные позы, одежда, ракурсы, освещение, сложный фон
- Прямое применение иллюстрированной структуры приводит к плохим результатам

http://www.robots.ox.ac.uk/~vgg/data/buffy_pose_classes/index.html

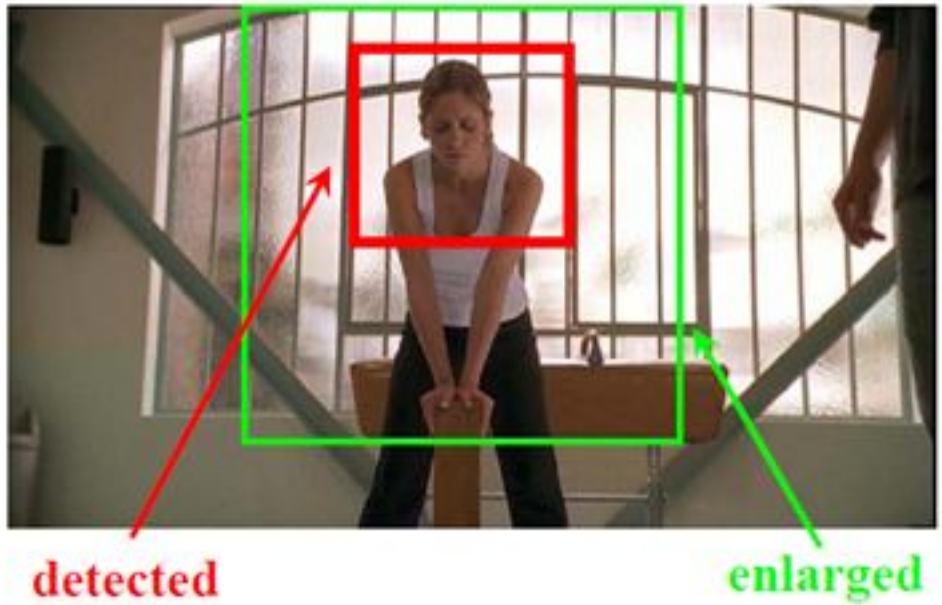
Детектор торса



- Использование детектора торса позволит нам локализовать область поиска
  - Задаёт масштаб тела и конечностей
  - Задаёт ограничение на положение конечностей
  - Работает быстро

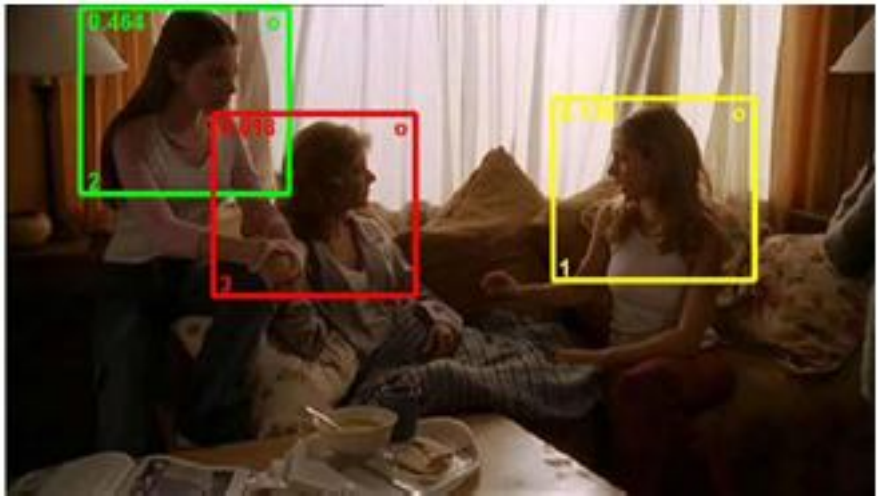
Инициализация по детектору

- HOG + SVM детектор, обученный по схеме bootstrapping



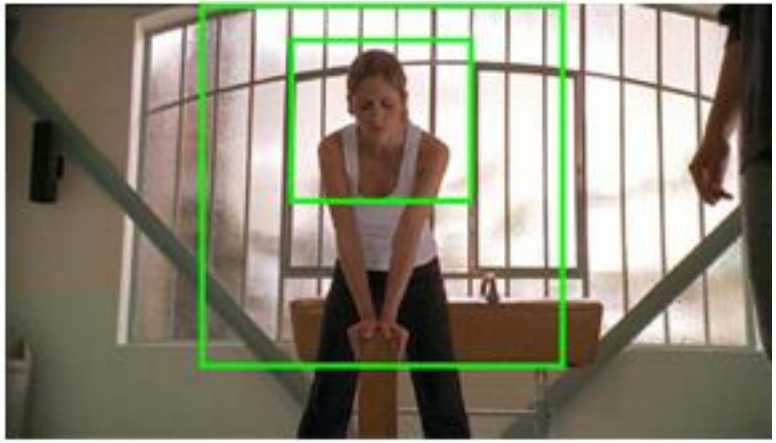
- Будем анализировать расширенную область торса, в которую должны попасть конечности

Детектор торса



- среднее изображение торса

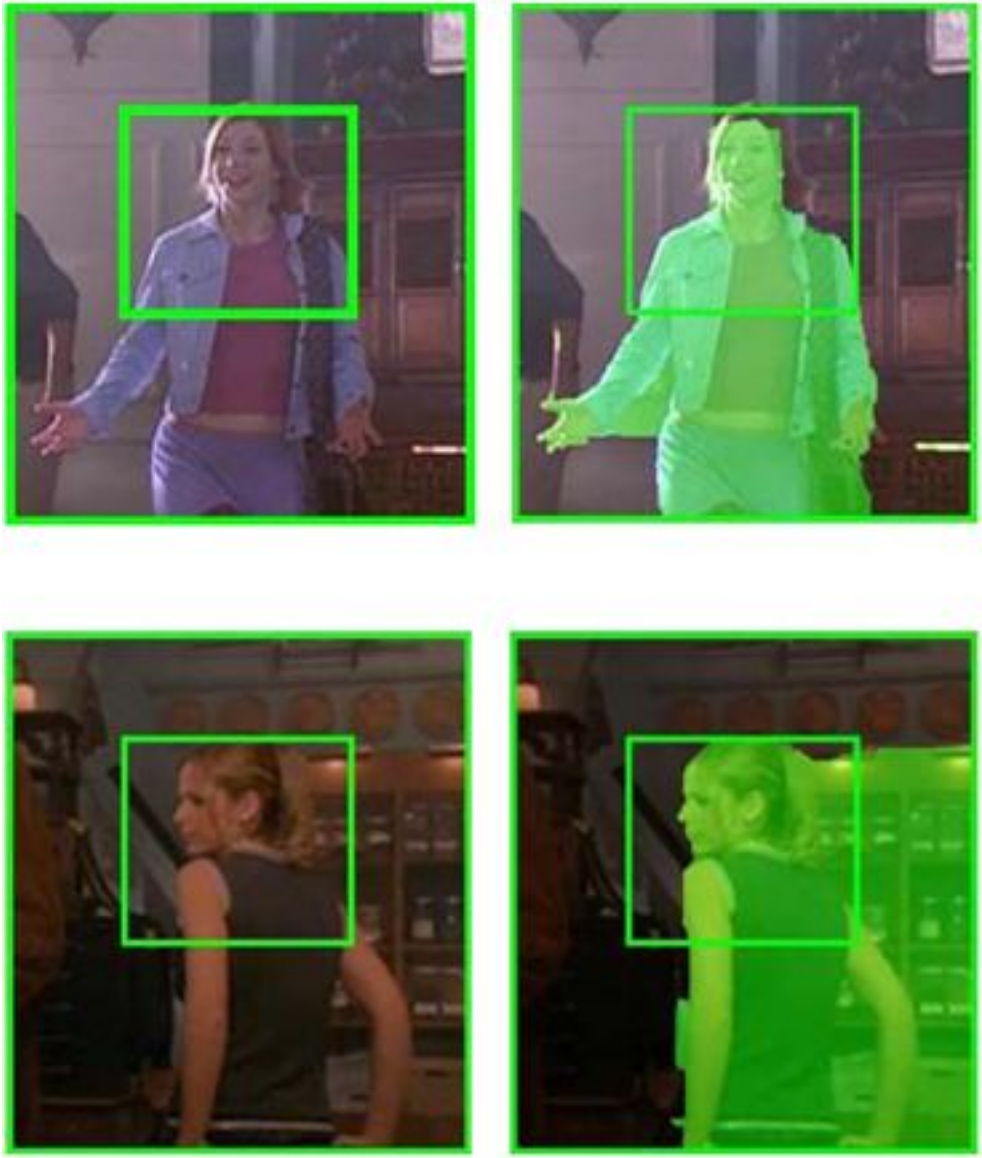
Выделение переднего плана



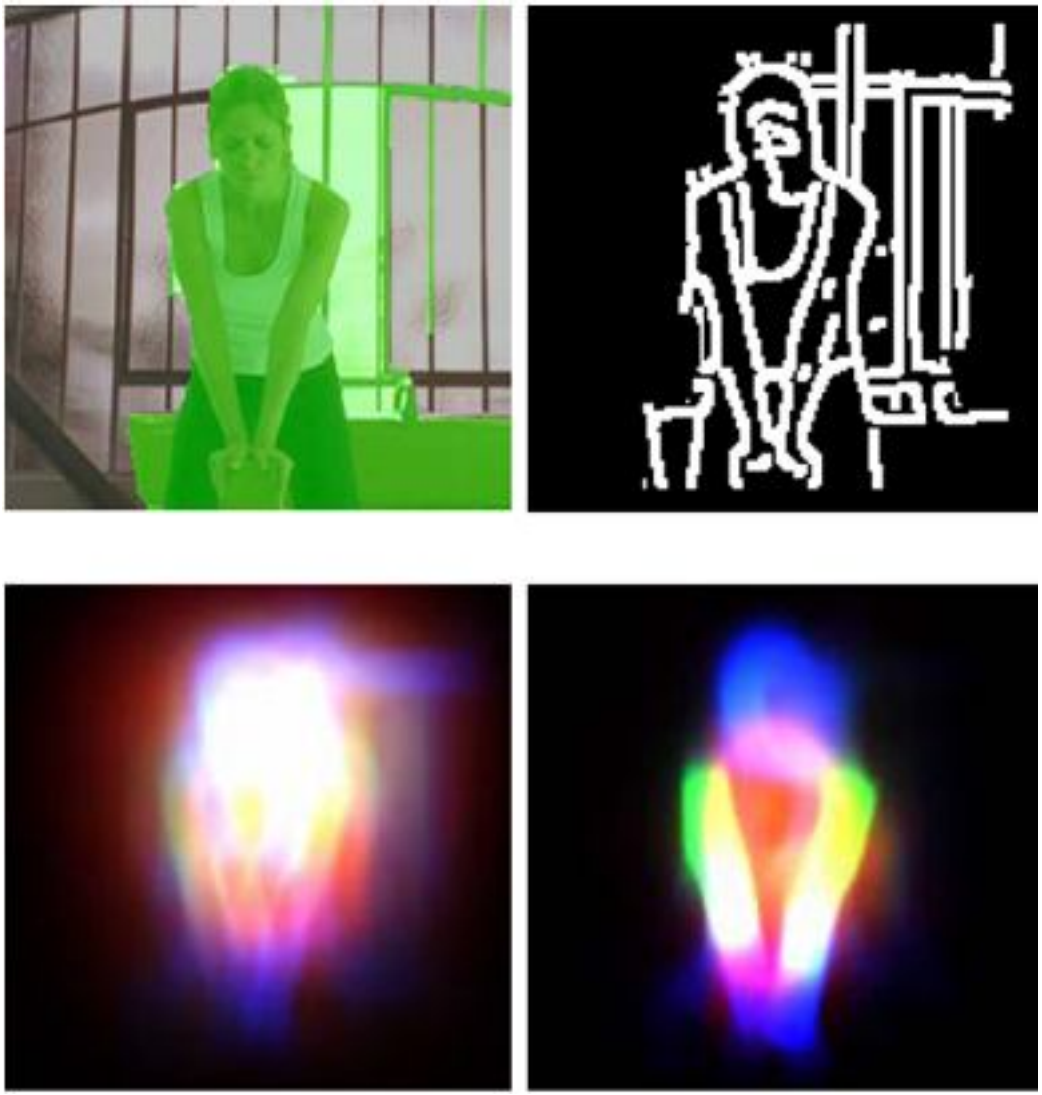
- Использование GrabCut для уточнения области поиска
- Инициализируем модели по выделенному торсу

- Консервативное уточнение (в 95.5% случаев человека не теряем)
- Уменьшаем область поиска
- Не требуется информации о фоне
- Фон может быть динамическим

Примеры



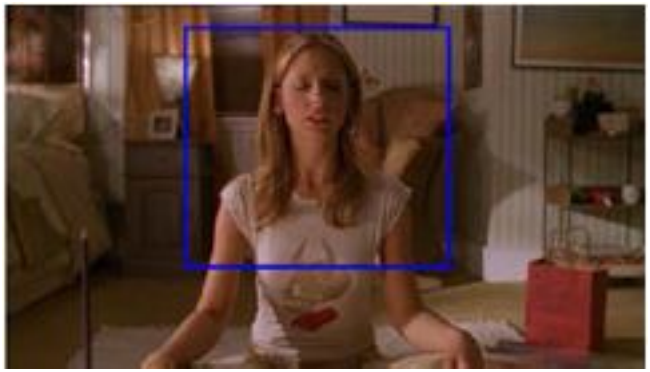
Вычисление позы

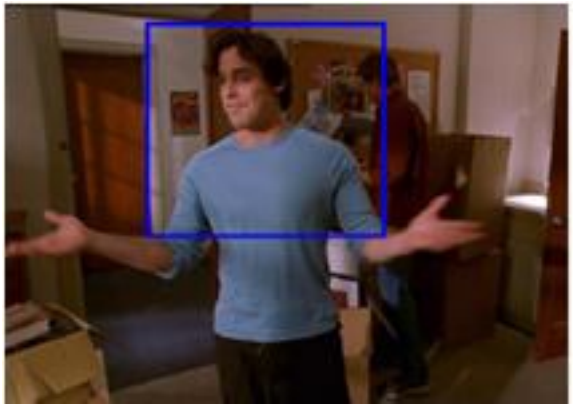


- Для оценки позы применяем иллюстрированную модель
  - Поиск краевыми шаблонами
  - Обучение модели цвета
  - Уточнение области поиска
- Итог:
  - Прогрессивное уточнение повышает скорость в 10-100 раз
  - Гораздо надежнее

Развитие

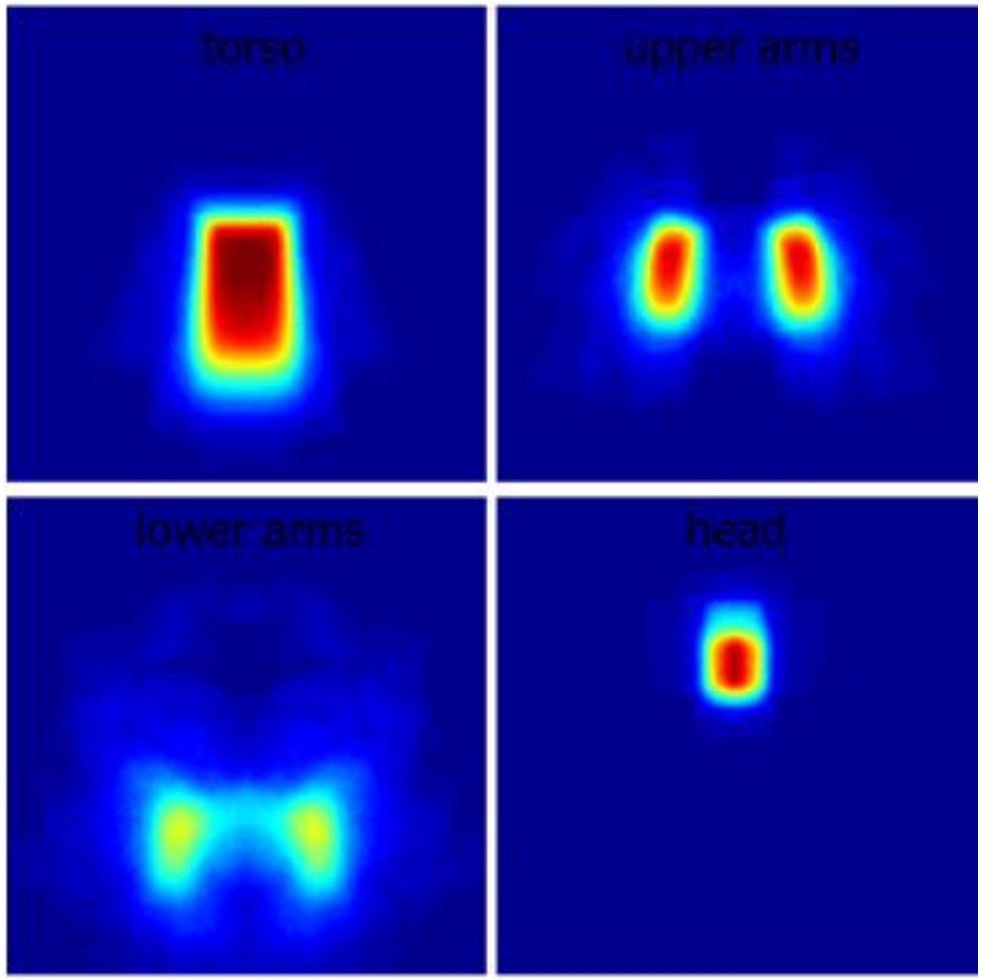
- Наблюдение 1:
  - Положение головы и торса стабильно относительно найденного торса
  - Плечи более подвижны относительно торса
  - Локти ещё более подвижны





Априорное расположение

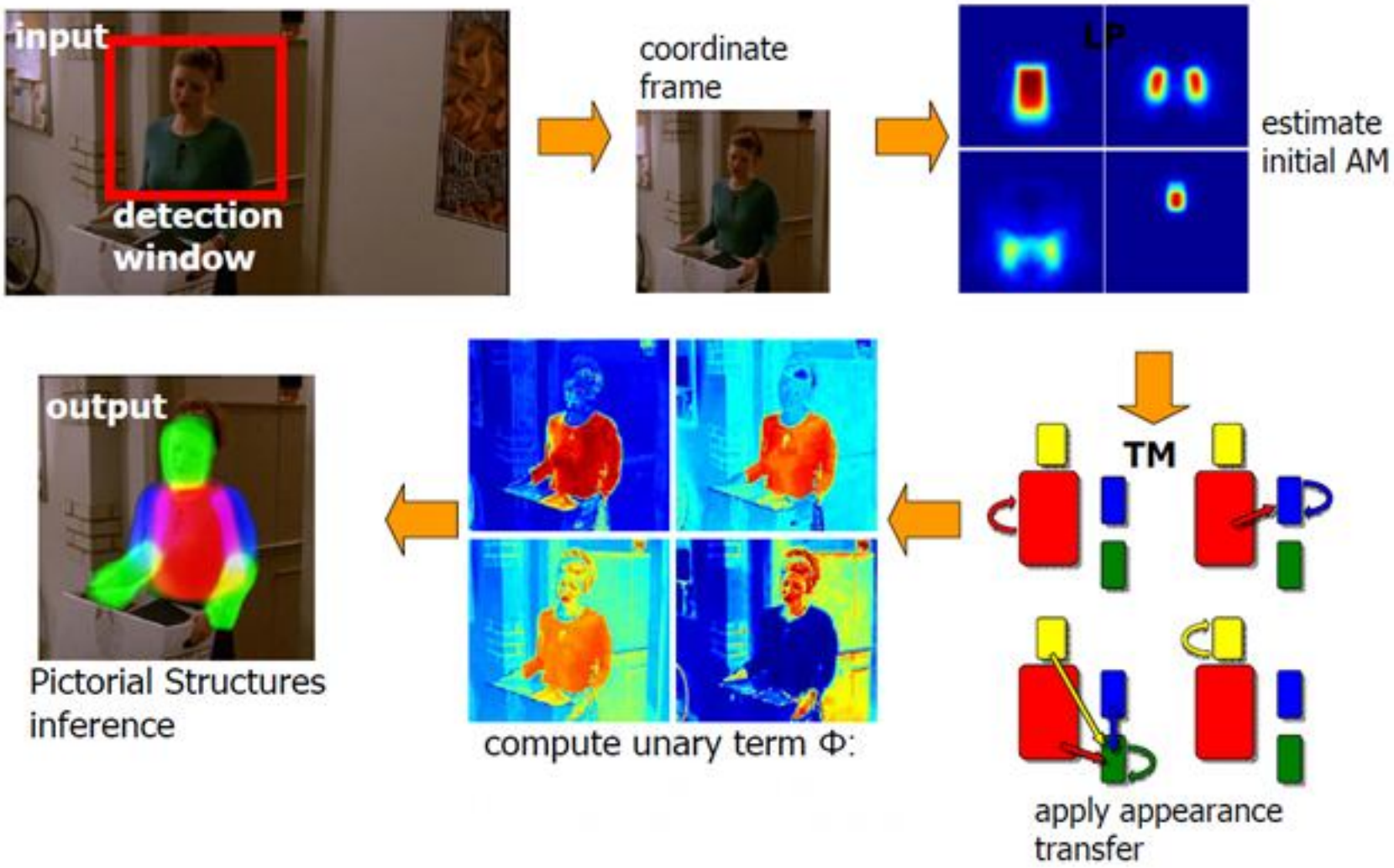
По выборке построим карты распределения положения элементов модели тела



Развитие

- Есть несколько стандартных шаблонов сходства цвета и текстуры отдельных частей модели
- Можем использовать внешность стабильных элементов модели для поиска подвижных частей

Расширенная схема целиком



Результаты

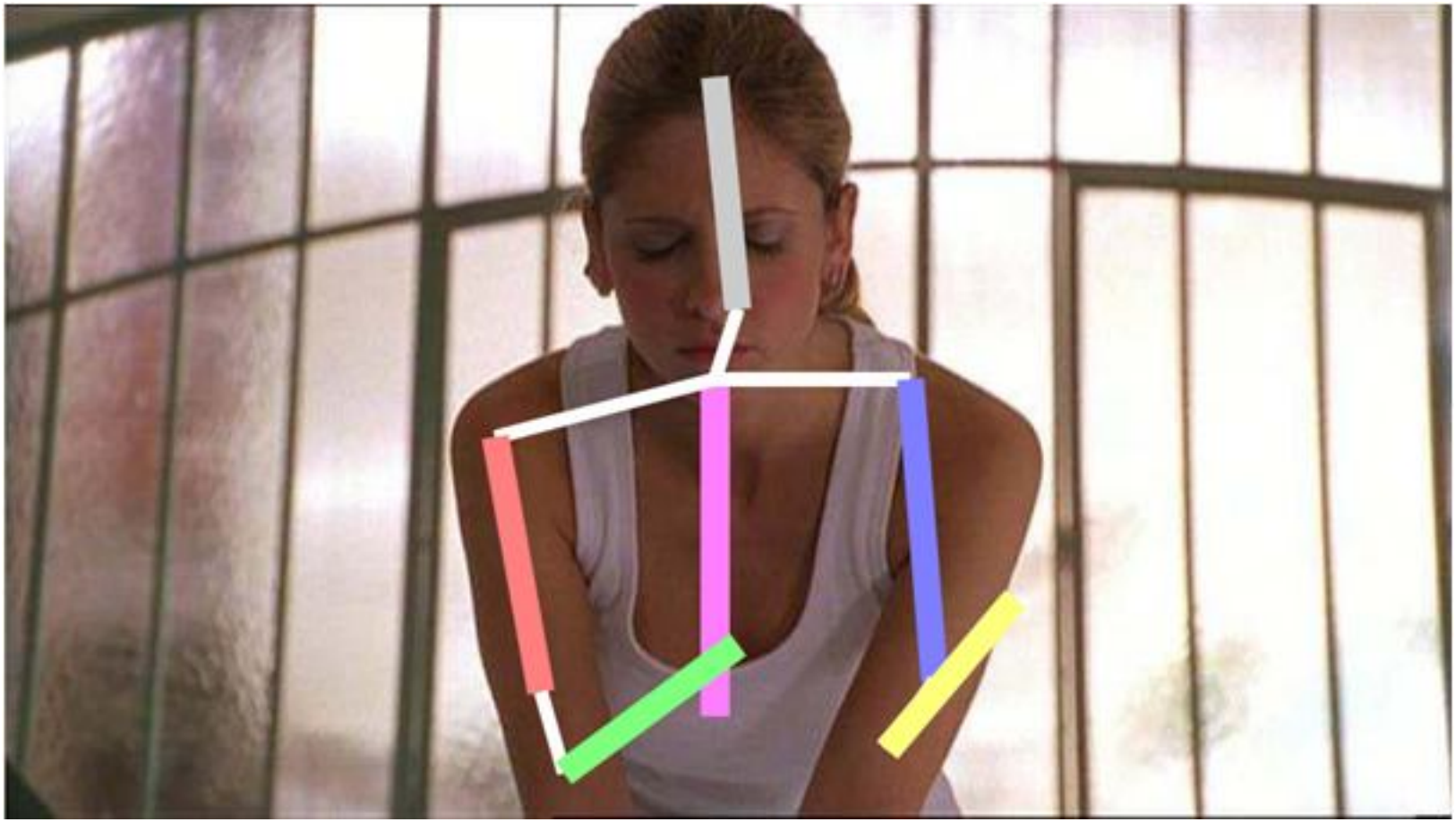
Качество демонстрируется на сложных сценах из Buffy-выборки

Ещё примеры

Ограничения моделей

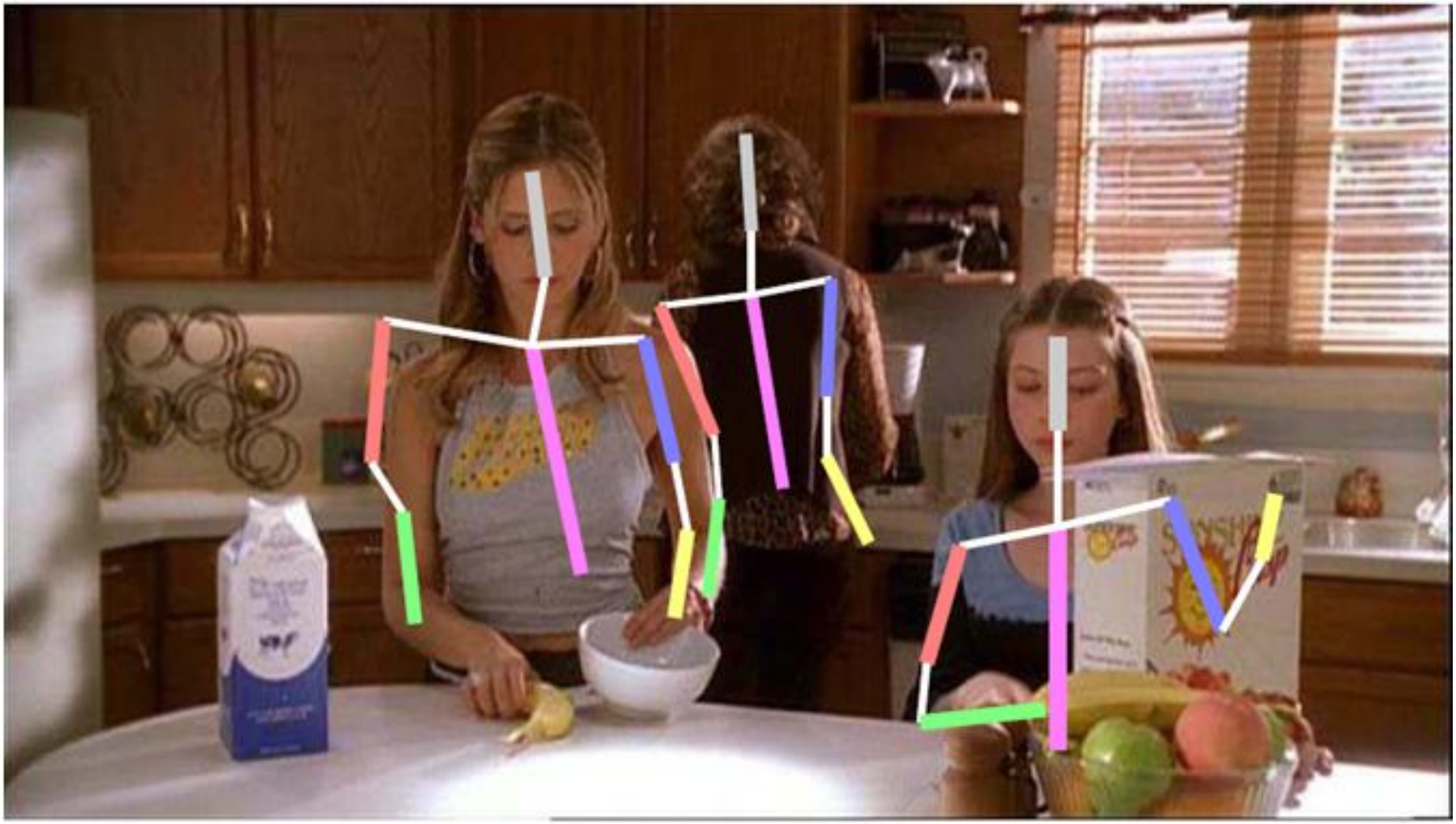
Когда все элементы видны и не перекрываются – модель применима

Обрезание кадра



Для части нет параметра видимости

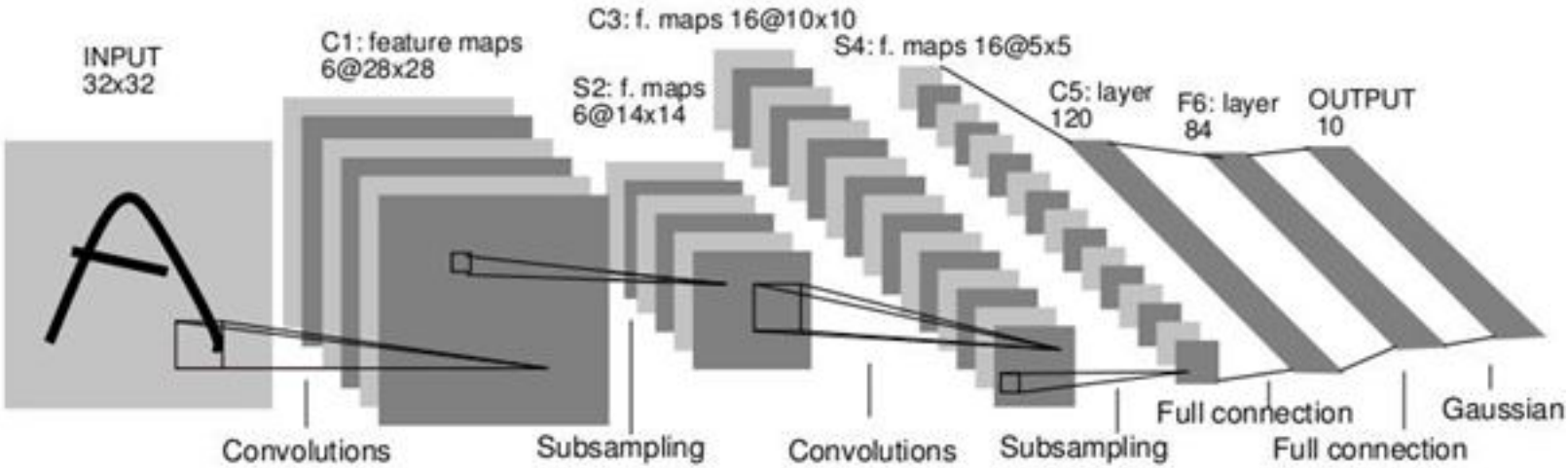
Перекрытия

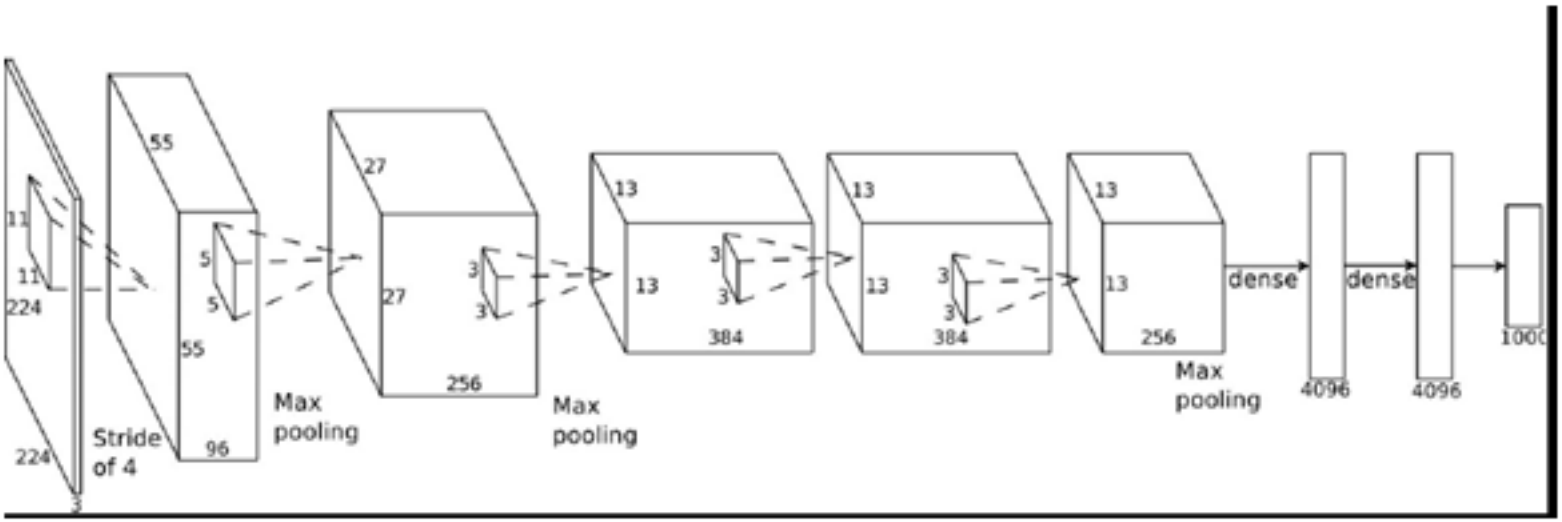


Резюме

- Существенную часть изменчивости изображений объектов можно смоделировать изменением положения частей относительно друг друга
- Есть целый ряд подходов к моделированию взаимного расположения частей и их использованию для поиска объектов
  - Многомерное нормальное распределение + PCA
  - Графические модели

Deep Learning





Pandemonium (1959)

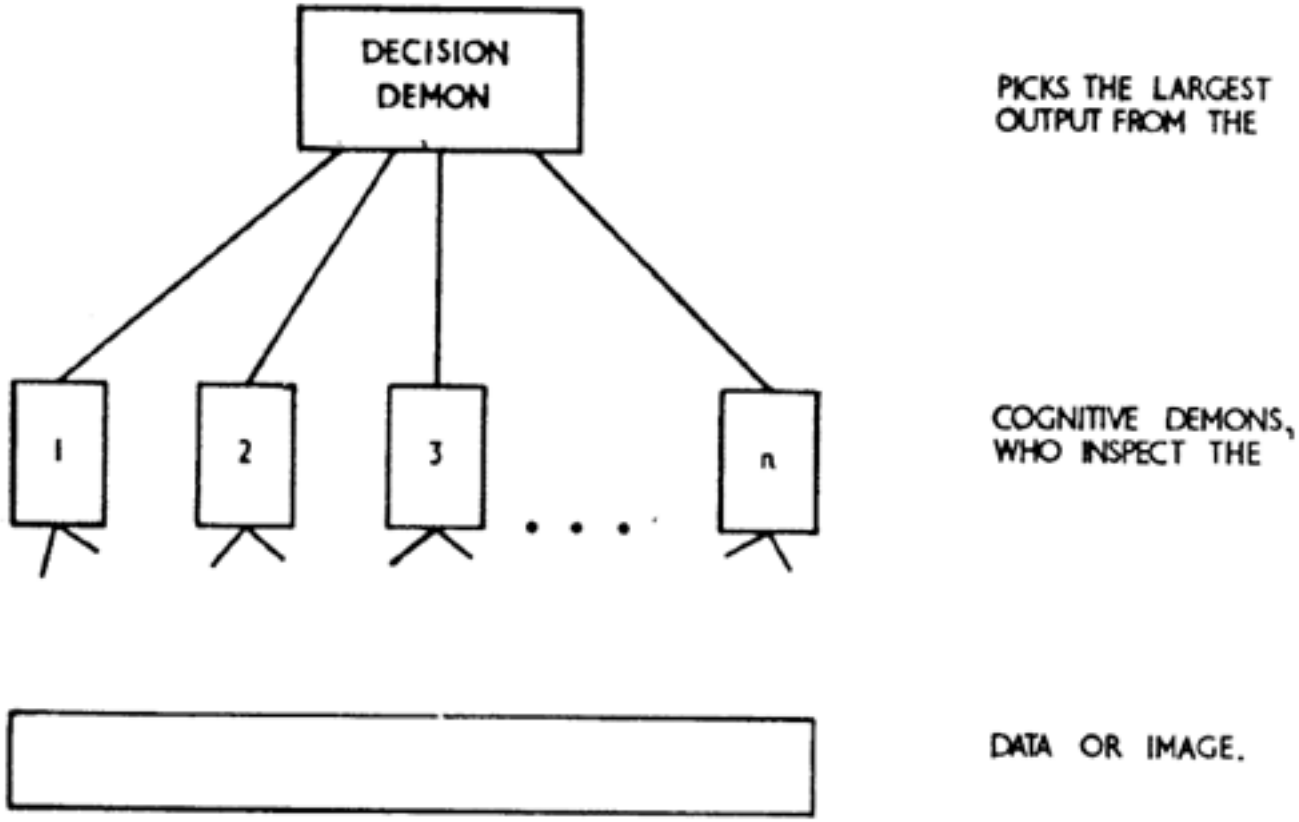
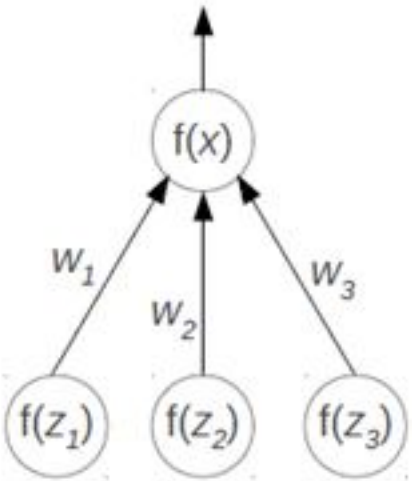


Fig. 1. Idealized Pandemonium.

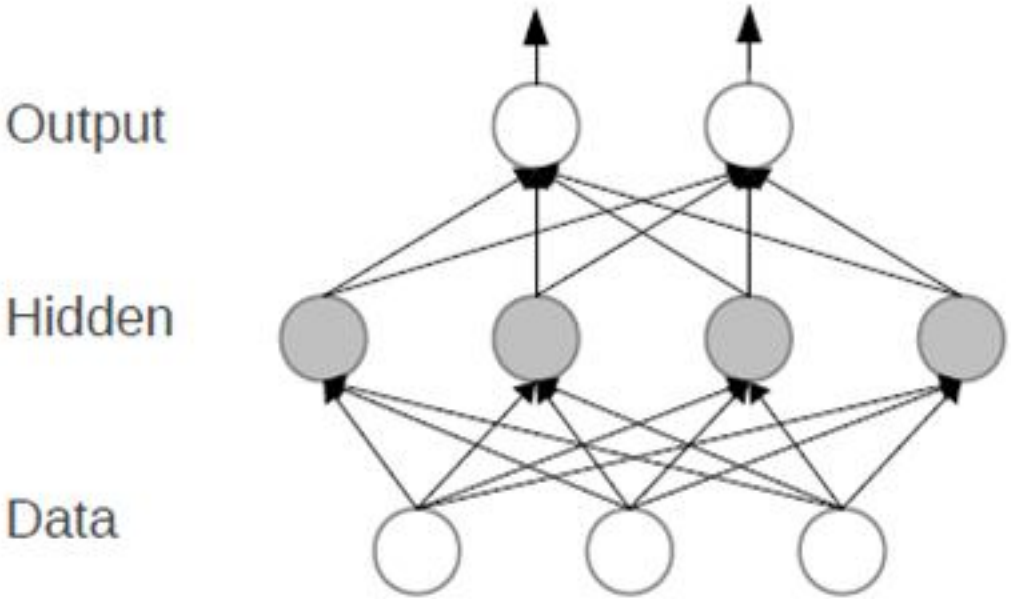
O.G. Selfridge. Pandemonium: a paradigm for learning. 1959

Нейрон и нейросеть



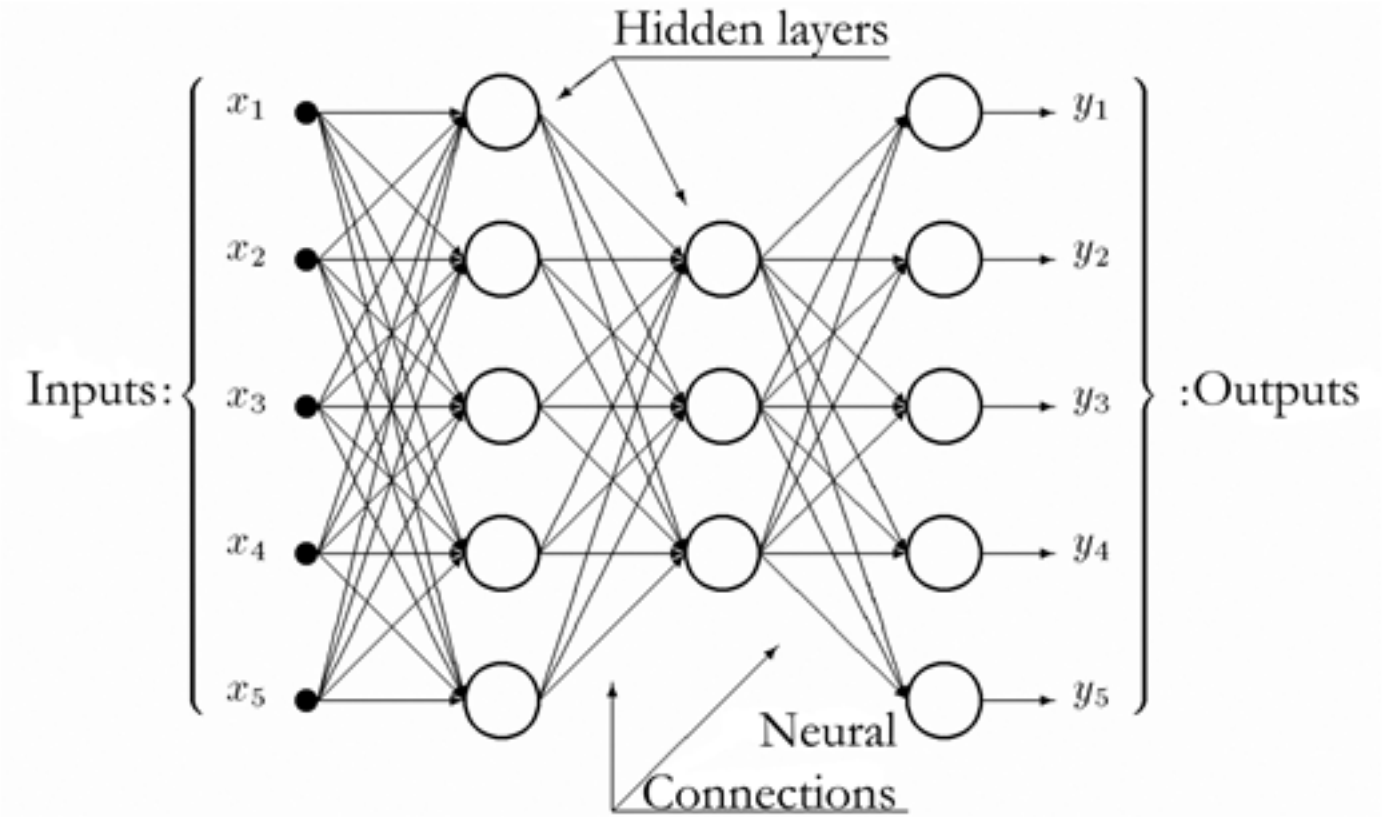
$x = w_1 f(z_1) + w_2 f(z_2) + w_3 f(z_3)$

x is called the total input to the neuron, and $f(x)$ is its output



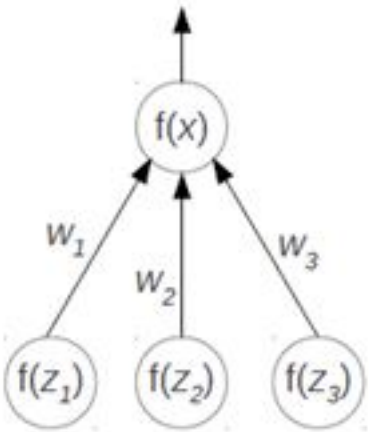
- У каждой связи есть «вес» – параметр
- Количество нейронов и количество связей определяет количество параметров

Feed-forward architectures



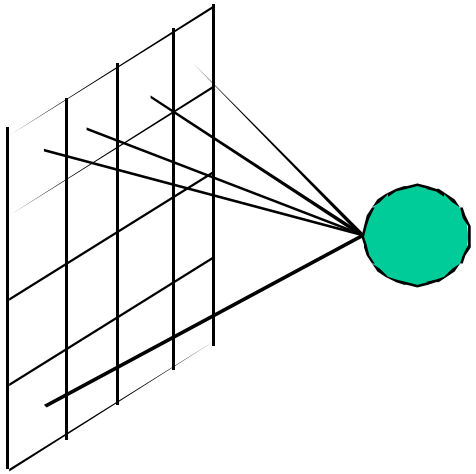
- Нейросети с обратными связями – рекуррентные (о них умолчим)
- Будем рассматривать только однонаправленные архитектуры
- Огромное количество параметров
  - Пример – полносвязанный слой

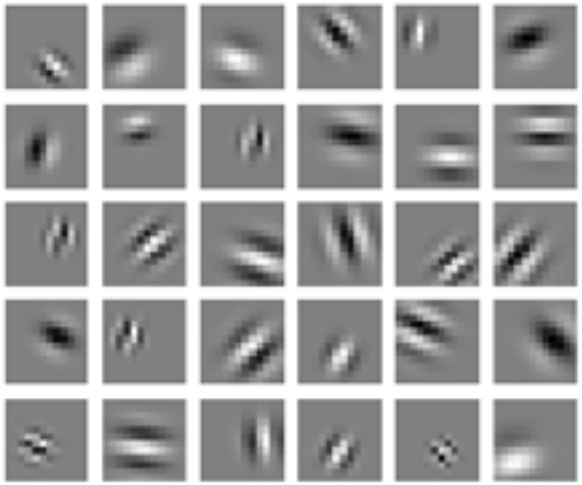
Нейрон как фильтр/признак



$x = w_1 f(z_1) + w_2 f(z_2) + w_3 f(z_3)$

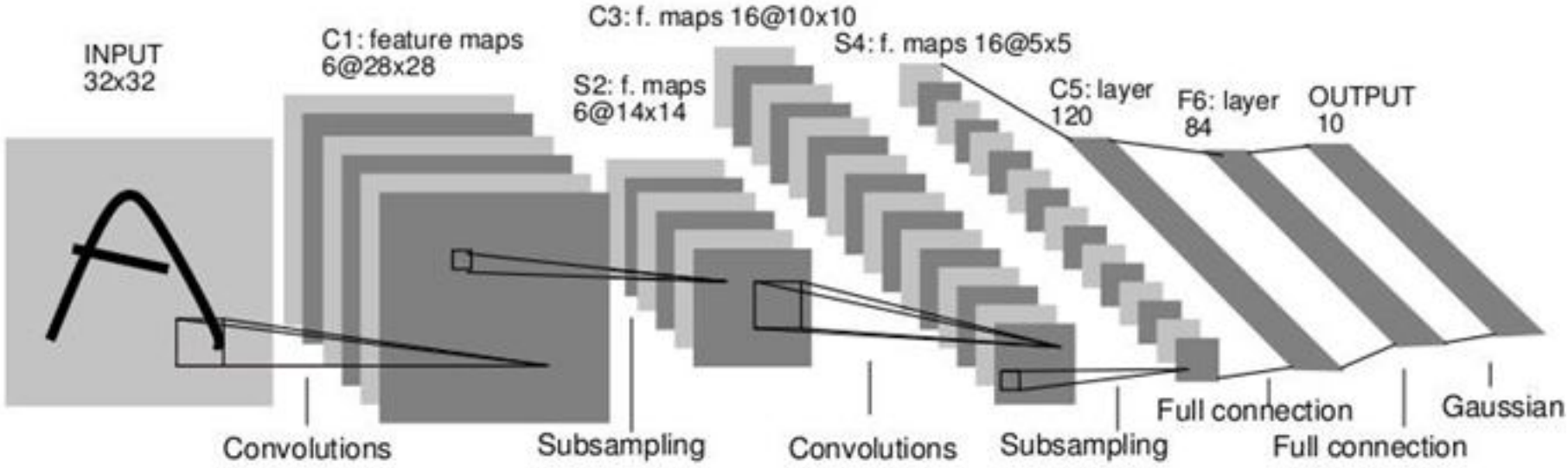
x is called the total input to the neuron, and $f(x)$ is its output





Операцию линейной (и не только) фильтрации одного пикселя (по окрестности) можно рассматривать как операцию, реализуемую нейроном

Свёрточные сети



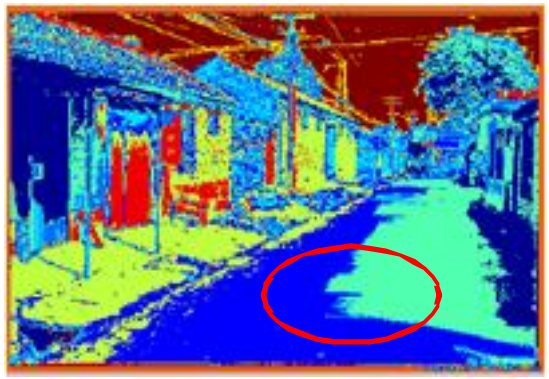
- «Свёрточный слой» реализует операцию свёртки
  - Каждой области одного размера на предыдущем слое соответствует свой нейрон
  - Вектор весов у всех нейронов одного слоя одинаковый
- Это резко снижает число признаков

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

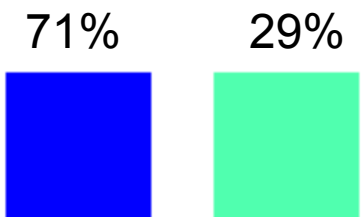
Общая схема вычисления признаков

Вычислим признаки по изображению → Квантуем признаки → Определим область для вычисления → Агрегируем признаки

RGB



Квантование на 10 уровней



Посчитанная гистограмма

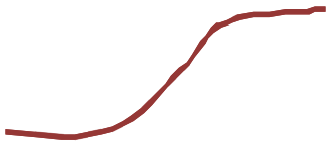
Features → Coding → Spatial pooling

Общая схема вычисления признаков

Пикселы / Признаки

Фильтрация по словарю

+ Нелинейность



Нормализация откликов

Группировка

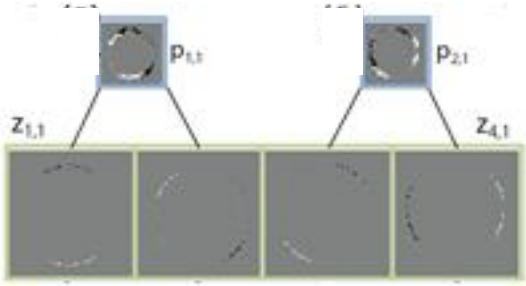
Разреженность

Max / Softmax

Нормализация локального контраста

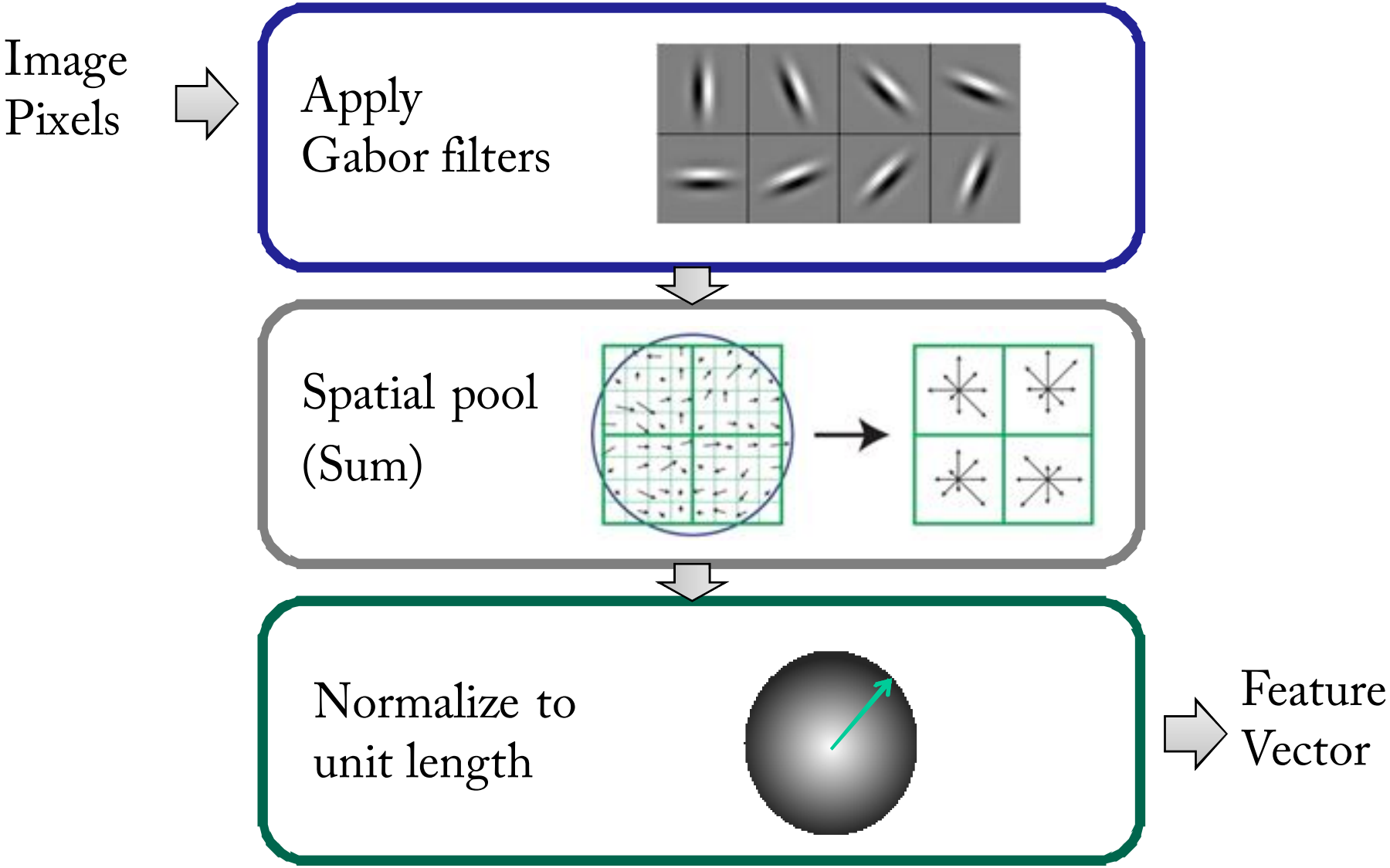
Нежесткое разделение

Аггрегация Области/признаки (Sum или Max)



Признаки

Дескриптор SIFT



Мешок слов по пирамиде

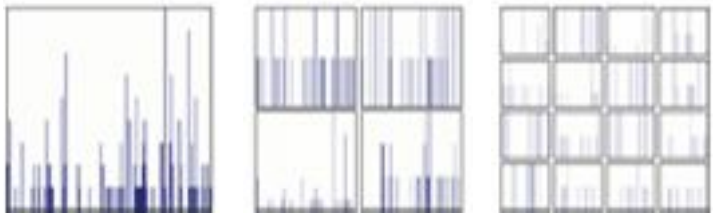
Lazebnik, Schmid, Ponce [CVPR 2006]

SIFT

Filter with Visual Words

Max

Multi-scale spatial pool (Sum)

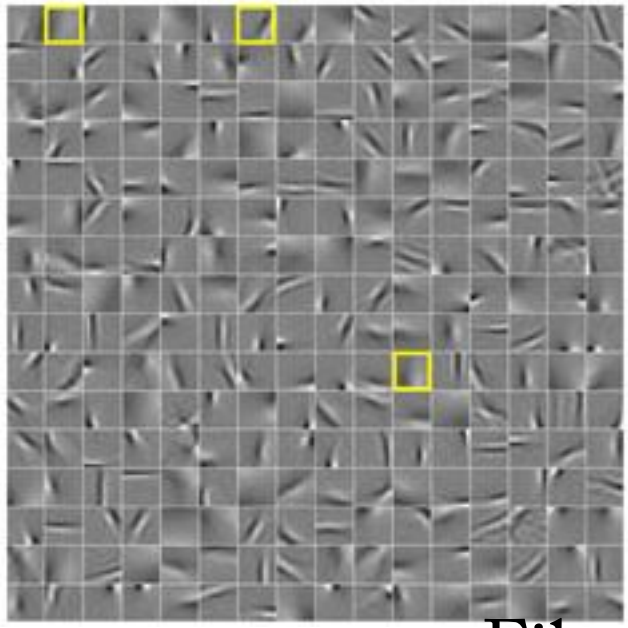


Classifier

Фильтрация

- Фрагменты
  - Изображение как набор фрагментов

Input

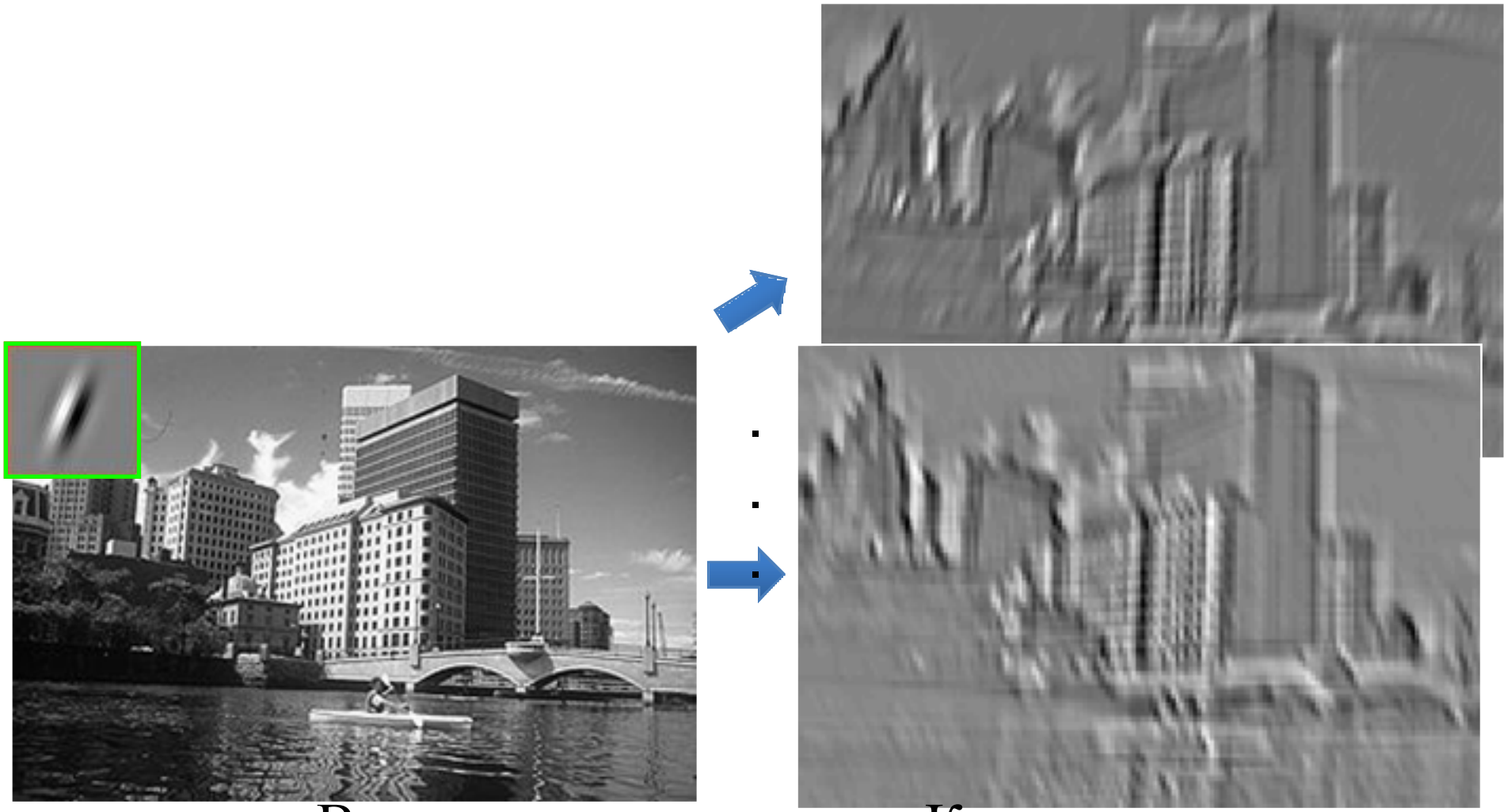


Filters

#patches

#filters

Фильтрация - свёртка

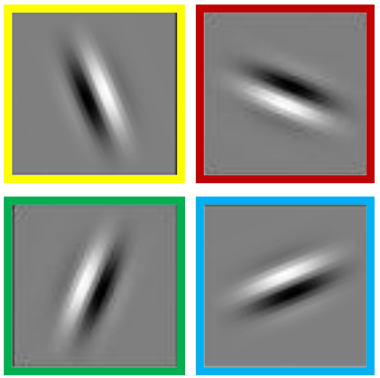


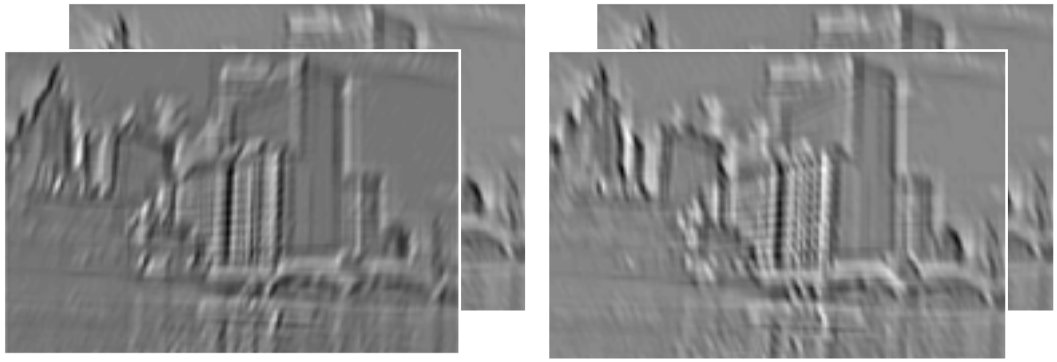
Ввод

Карты признаков

Фильтрация - тайлинг

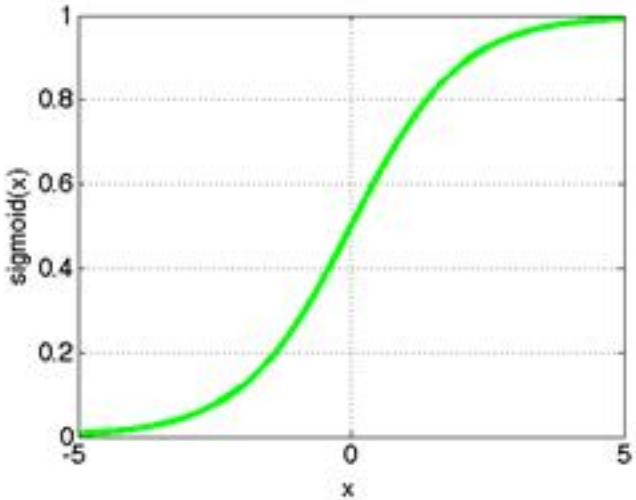
- Повторяем фильтры каждый n пикселей
- Можно использовать больше фильтров для того же числа признаков

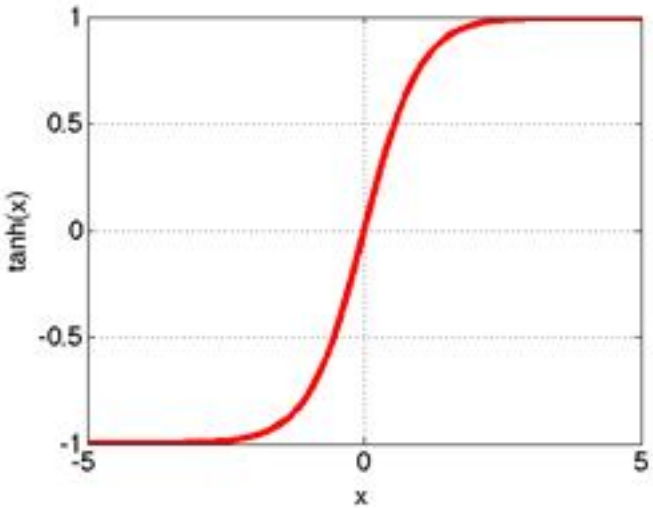


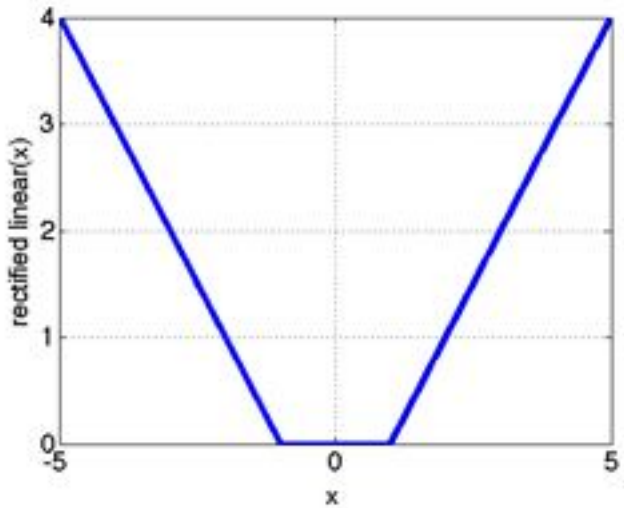


Фильтрация

- Нелинейность
  - Tanh
  - Sigmoid: 1/(1+exp(-x))
  - Rectified linear

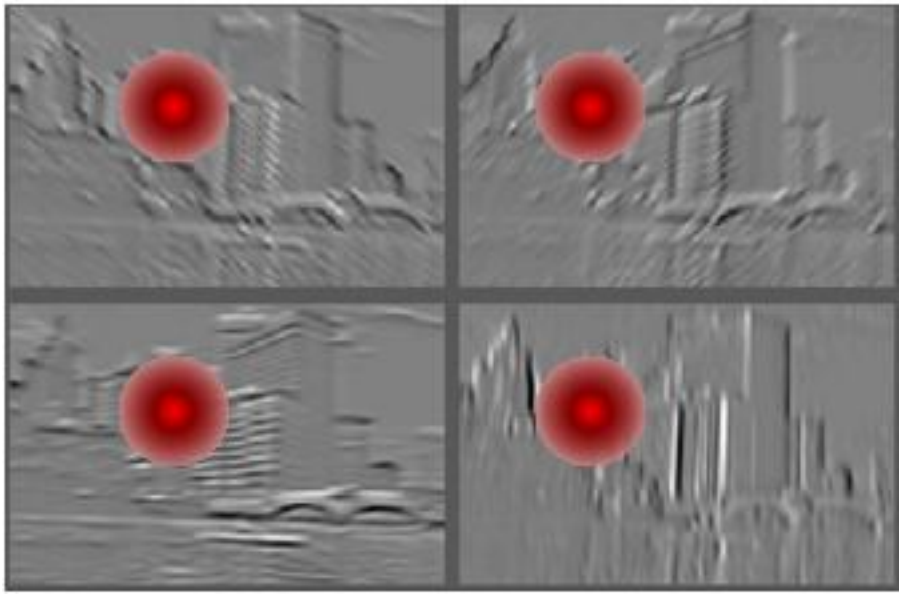




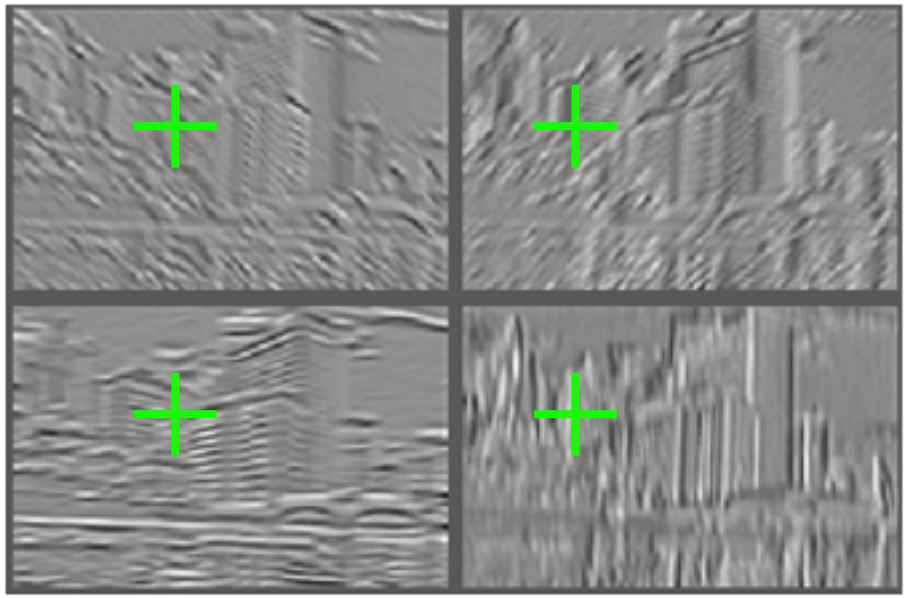


Нормализация контраста

- Local mean = 0, local std. = 1, “Local” → 7x7 Gaussian



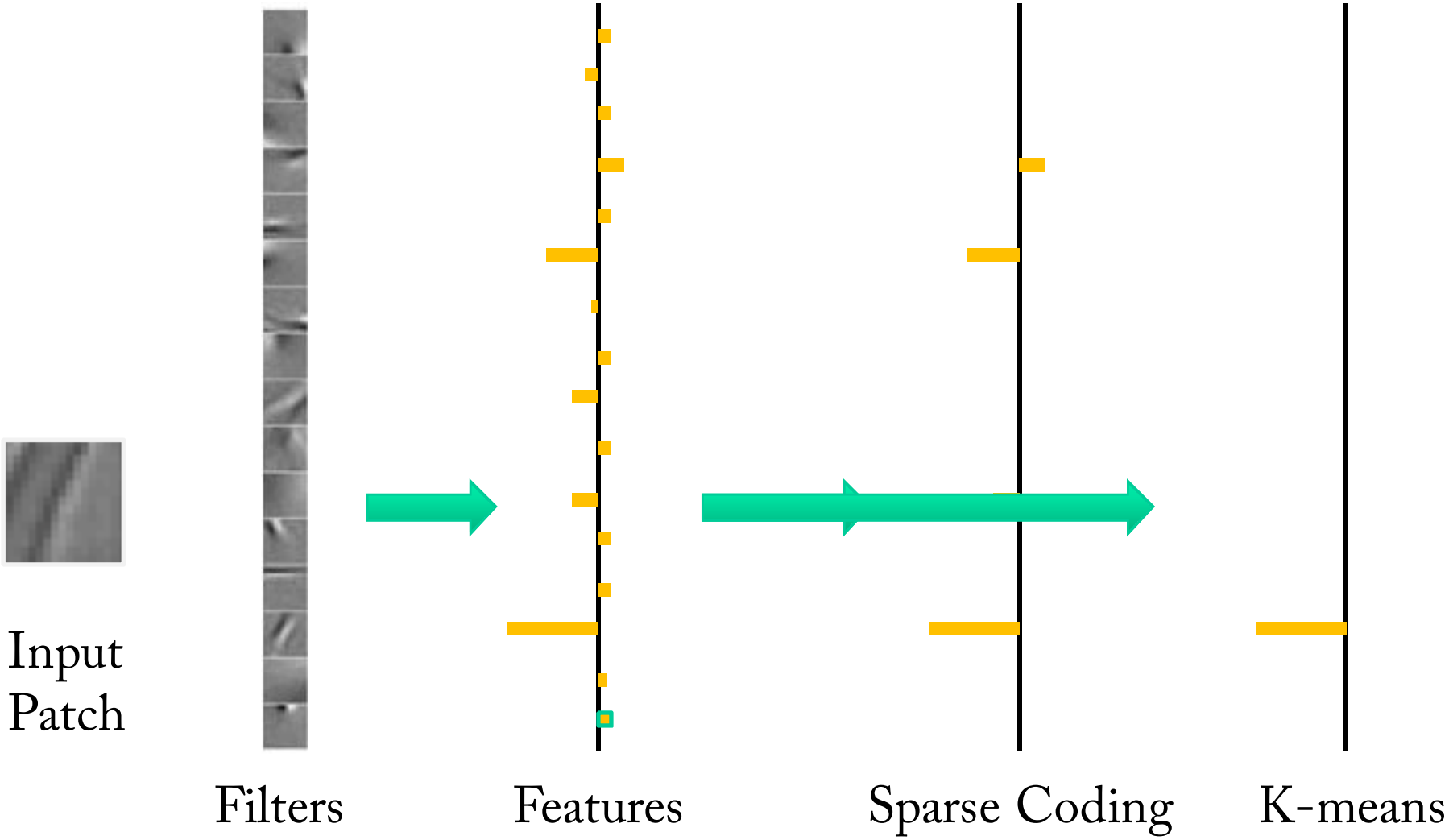
Карты признаков



После нормализации контраста

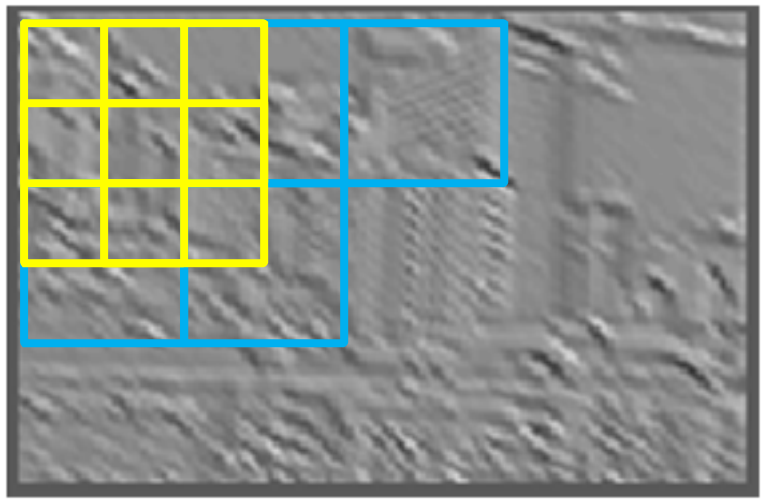
Нормализация

- Разреженность – ограничение по L_0 or L_1

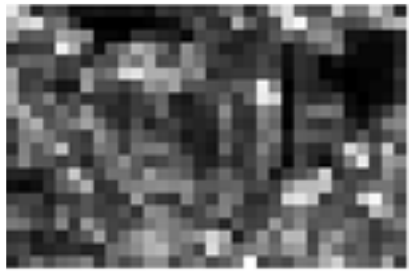


Пространственная агрегация

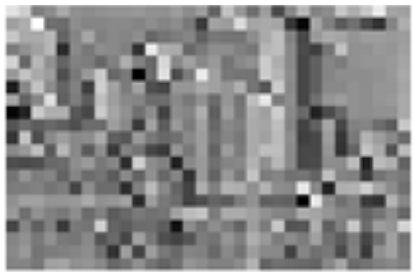
- Не перекрывающиеся / перекрывающиеся области
- Sum или max
- Boureau et al. ICML'10 – теоретический анализ



Max



Sum



Промежуточный вывод

- Многослойные направленные нейросети – система иерархического расчёта признаков
- «Конфигурация» сети – определение типа слоёв (свёрточный, полносвязанный, масштабирующий, нормализующий и т.д.)
- После определения конфигурации сети встаёт вопрос оценки параметров (весов) сети

Geofrey Hinton «To Recognize Shapes, First Learn to Generate Images», Technical report, 2006

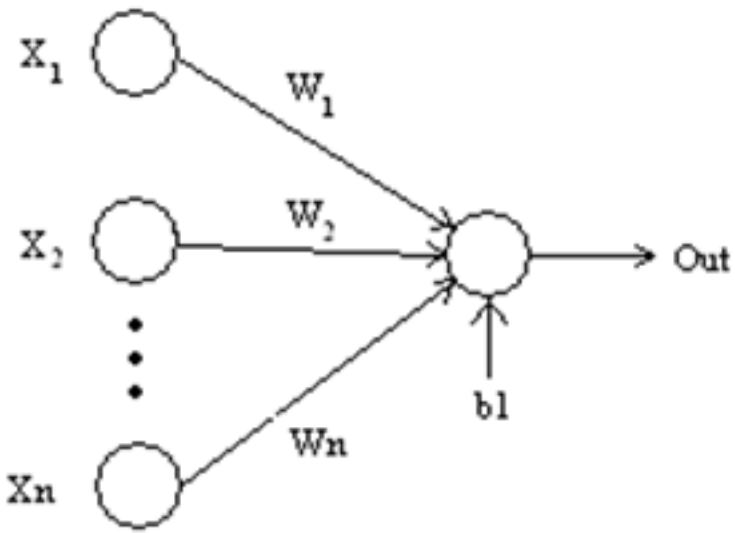
Стратегии обучения

- Отрицание проблемы (denial)
  - Однослойные сети
  - Эвристическое задание параметров
- Эволюционная стратегия
  - Случайно меняем (jitter) веса и оставляем, если сеть улучшилась
- Оптимизация всех параметров в совокупности за счёт матана (calculus)
  - Градиентный спуск – обратное распространение ошибки
- Прокрастинация
  - Откладываем главную задачу «на потом», обучаем веса на каждом слое так, чтобы они извлекали какую-то полезную информацию

Geofrey Hinton «To Recognize Shapes, First Learn to Generate Images», Technical report, 2006

Отрицание проблемы

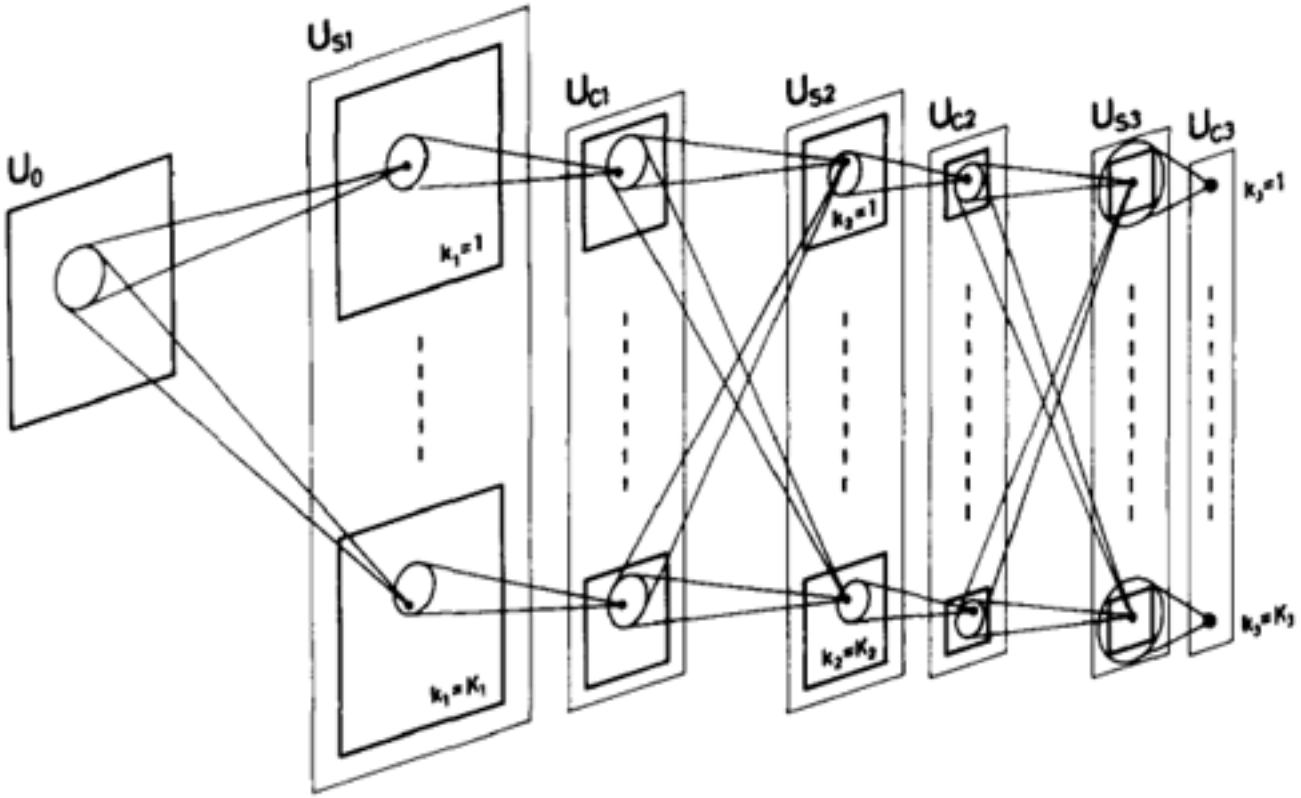
Однослойные персептроны



- Будем подбирать параметры только одного слоя, принимающего решения
- Вход xi – признаки данных
- Современный вариант персептрона - SVM

- Rosenblatt, F. (1962). Principles of neurodynamics. Spartan Books.
- Minsky, M. and Papert, S. (1969). Perceptrons: an introduction to computational geometry. MIT Press Cambridge, Mass.

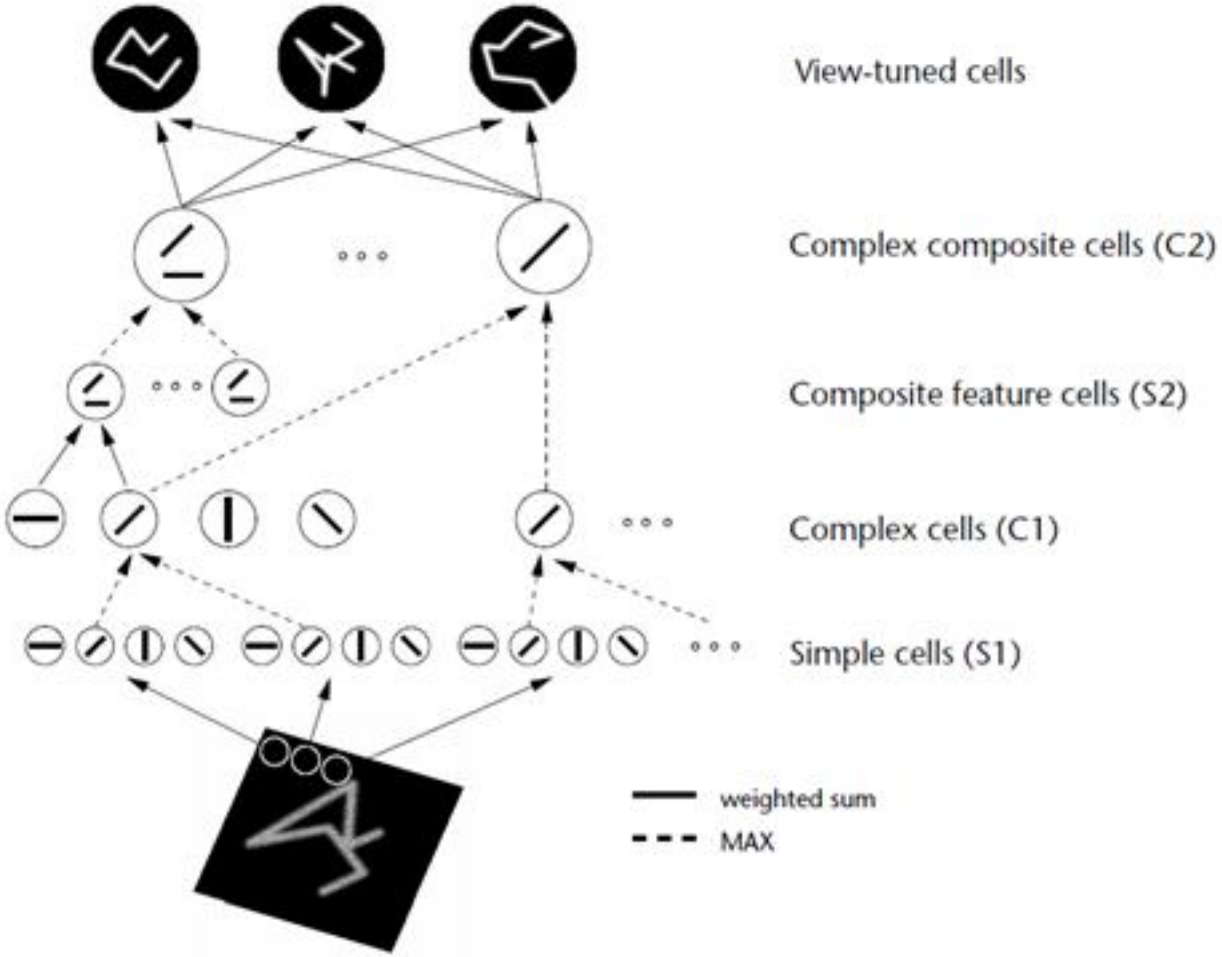
Neucognitron (1980)



- S-слои – линейные фильтры изображения
- C-слои – MAX операторы, дающие инвариантность
- На верхнем уровне инвариантность по положению по всему изображению

K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36(4): 93-202, 1980.

Модель HMAX



Максимум по всему изображению

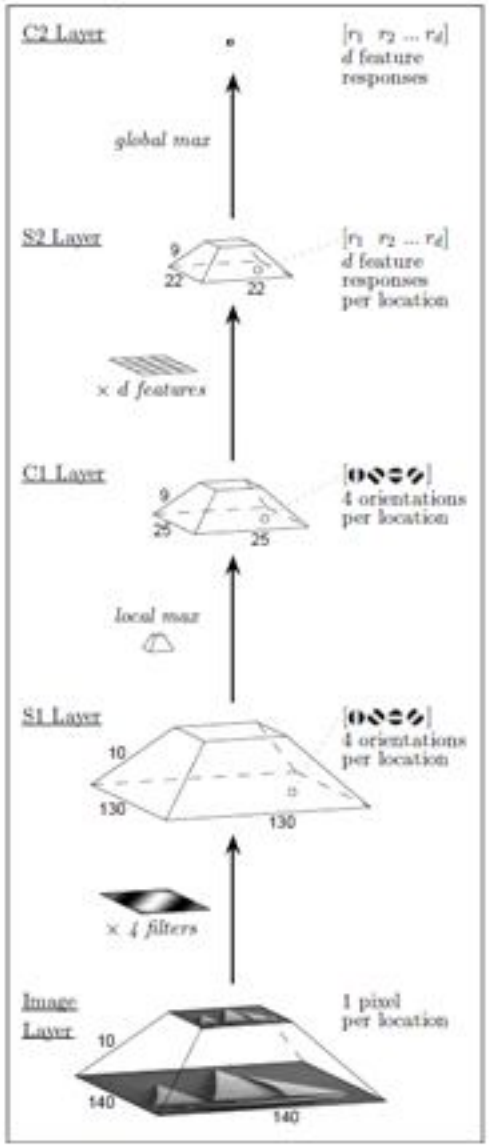
Фильтры-фрагменты

Максимум по окрестности

Фильтры Габора

Riesenhuber, M. and T. Poggio. Hierarchical Models of Object Recognition in Cortex, *Nature Neuroscience*, 2, 1019-1025, 1999.

Признаки в модели HMAX



Обучение:

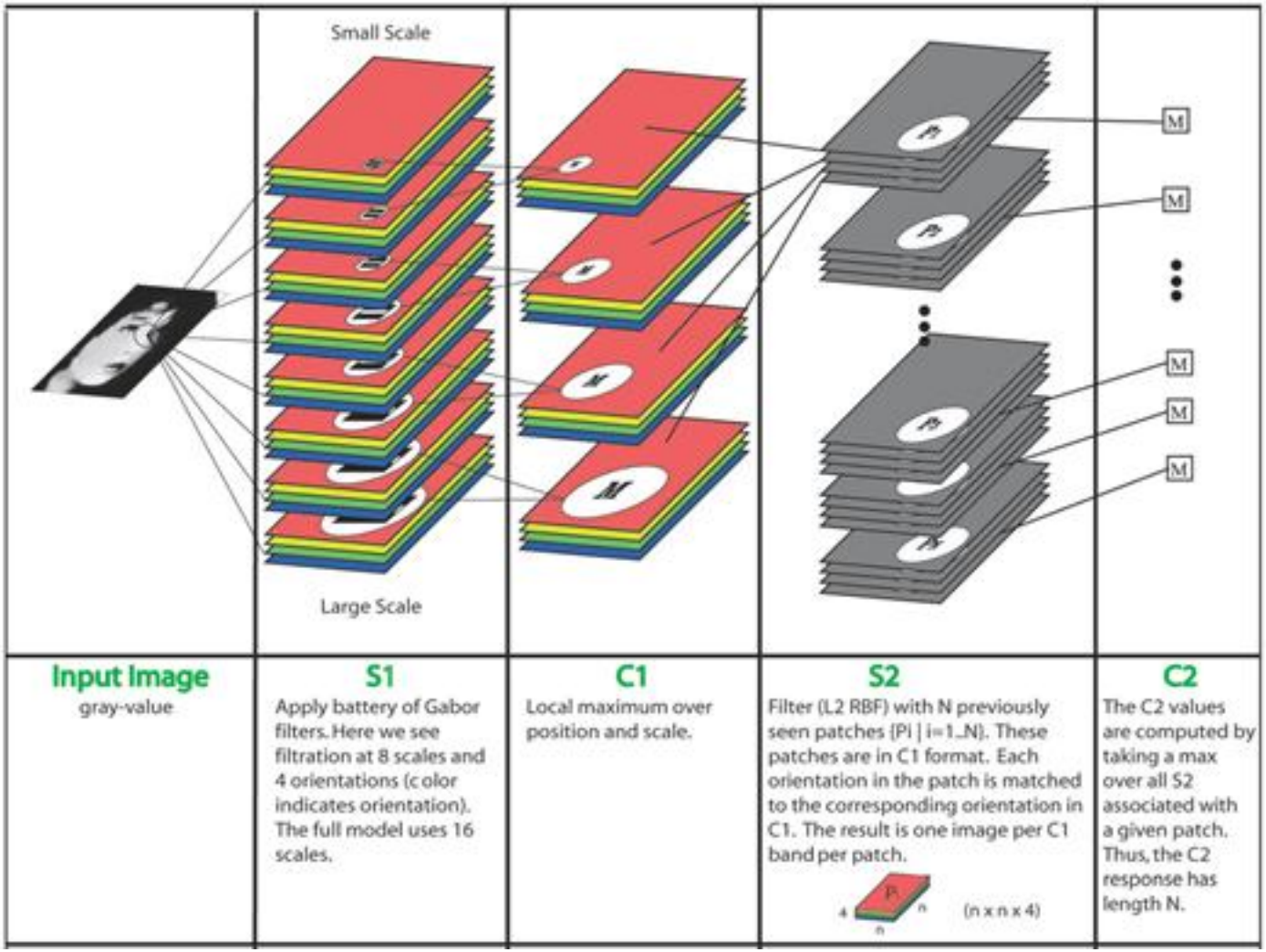
- Выбор N случайных фрагментов
- Возможно, проверка и выбор наиболее информативных фрагментов
- Каждый фрагмент – «признак» высокого уровня

Выход алгоритма:

- N переменных, показывающих максимальный отклик каждого из N высокоуровневых признаков
- Классификация через SVM и т.д.

Serre, T., L. Wolf and T. Poggio. Object Recognition with Features Inspired by Visual Cortex. CVPR 2005.

Другая визуализация



Распознавание событий

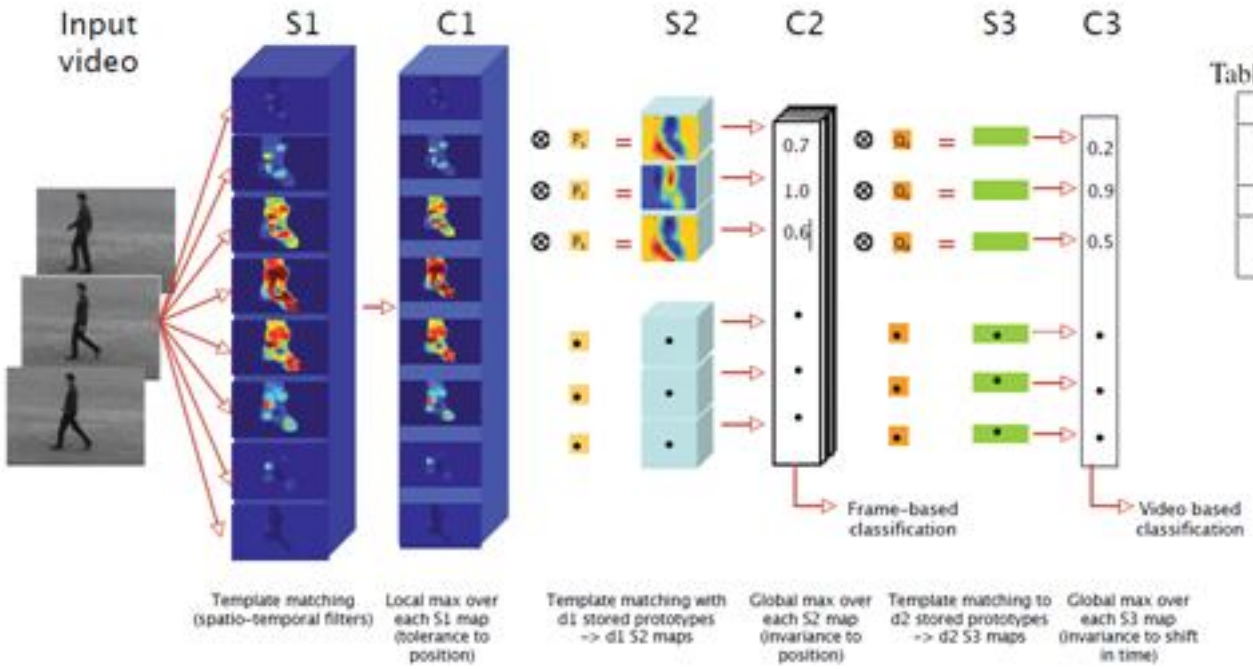


Table 6. Average performance for shape *vs.* motion cues.

HOG/HOF	HOGHOF	HOG	HOF
Original	**20.44**%	15.01%	17.95%
Stabilized	21.96%	15.47%	**22.48**%
C2	Motion+Shape	Shape	Motion
Original	**22.83**%	13.40%	21.96%
Stabilized	**23.18**%	13.44%	22.73%

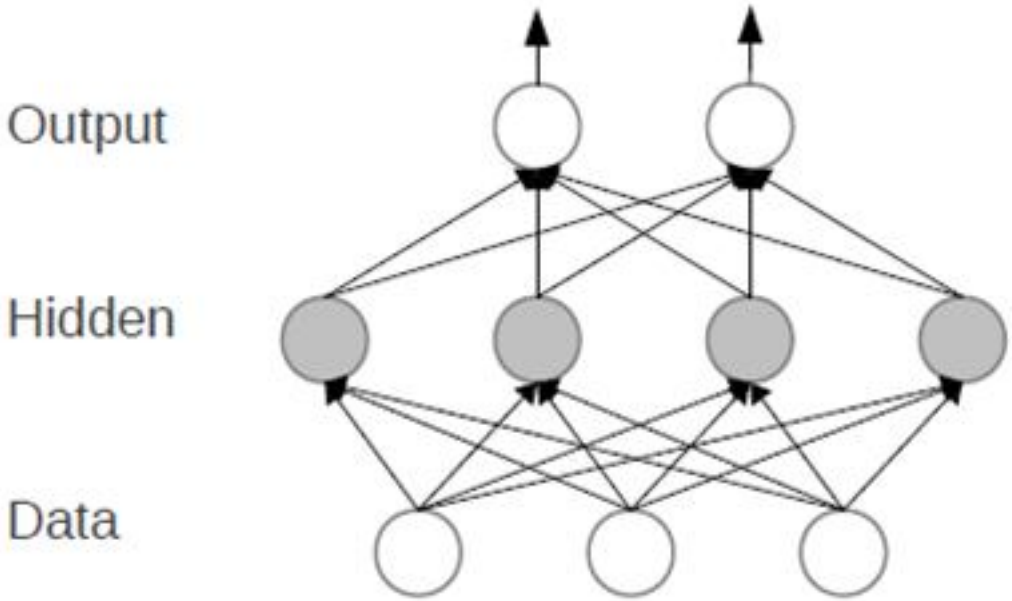
Точность на HMDB (2011) в сравнении HOG/HOF признаками

- Замена S1 признаков изображений на признаки по пространственно-временному объёму
- Например I_t/I_x и I_t/I_y

H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. ICCV, 2007

Градиентный спуск

Нейросеть



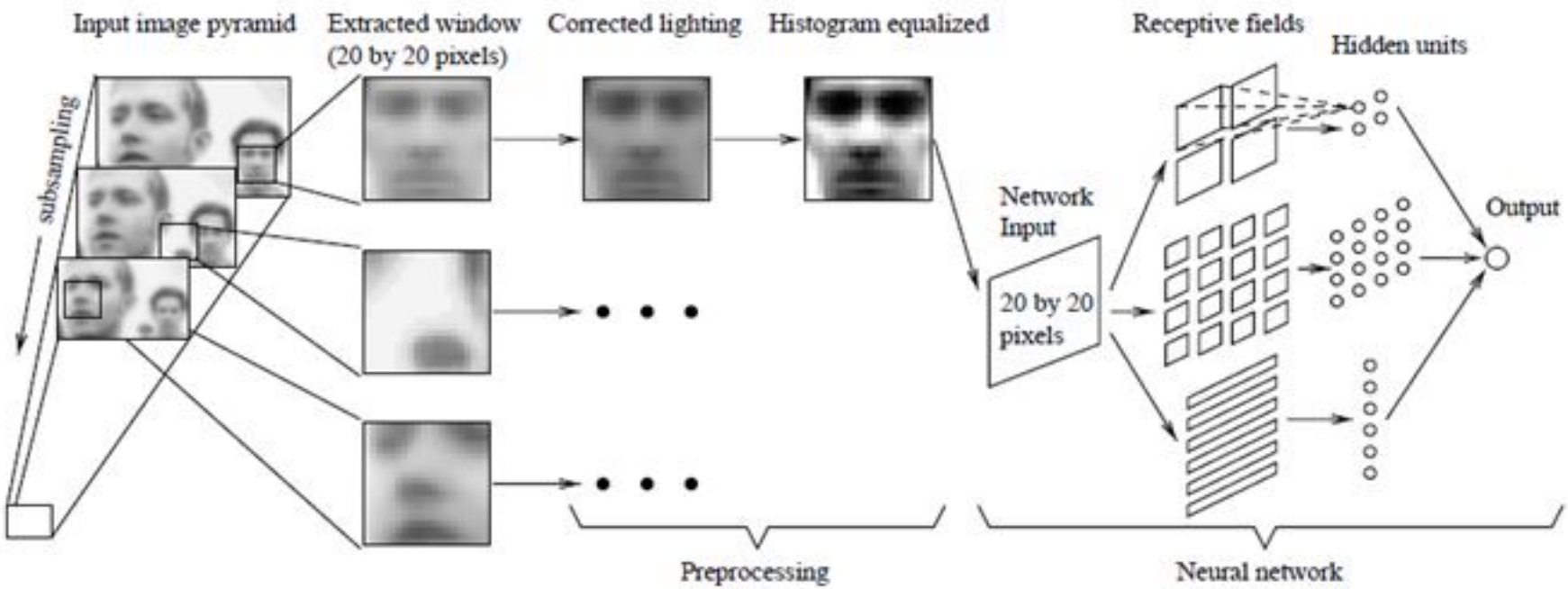
Нейросеть вычисляет дифференцируемую функцию от своих входов

•↑ *Галушкин А. И.* Синтез многослойных систем распознавания образов. — М.: «Энергия», 1974.

•↑ *Werbos P. J.*, Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph.D. thesis, Harvard University, Cambridge, MA, 1974.

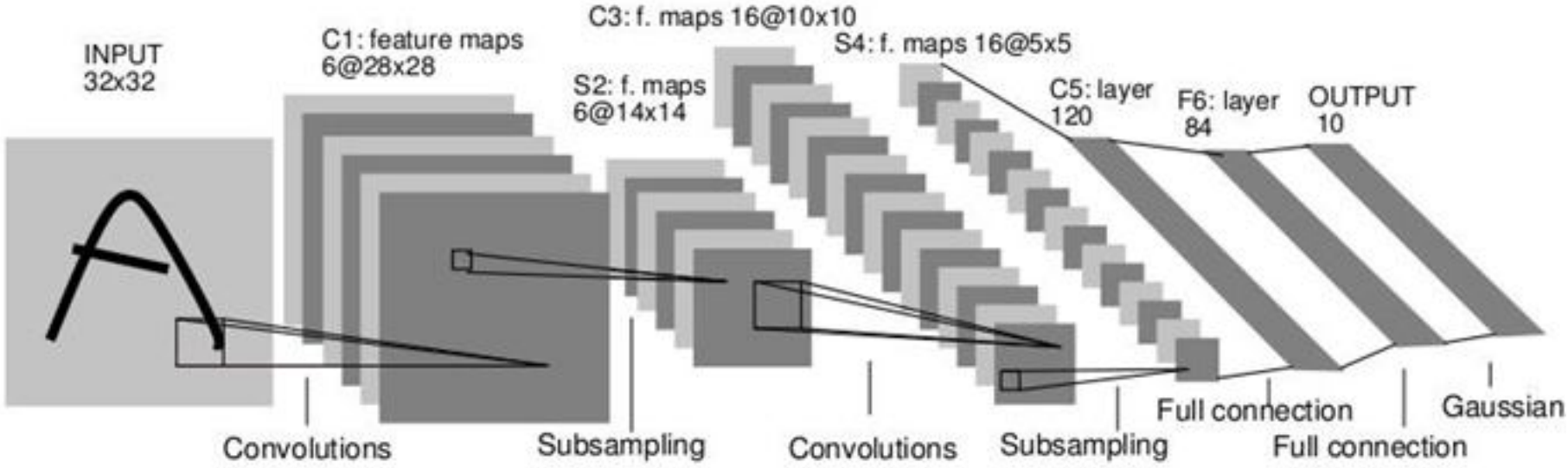
•↑ [1] [2] *Rumelhart D.E., Hinton G.E., Williams R.J.*, Learning Internal Representations by Error Propagation. In: Parallel Distributed Processing, vol. 1, pp. 318—362. Cambridge, MA, MIT Press. 1986.

Rowley face detector (1998)



Rowley, Baluja, and Kanade: Neural Network-Based Face Detection (PAMI, January 1998)

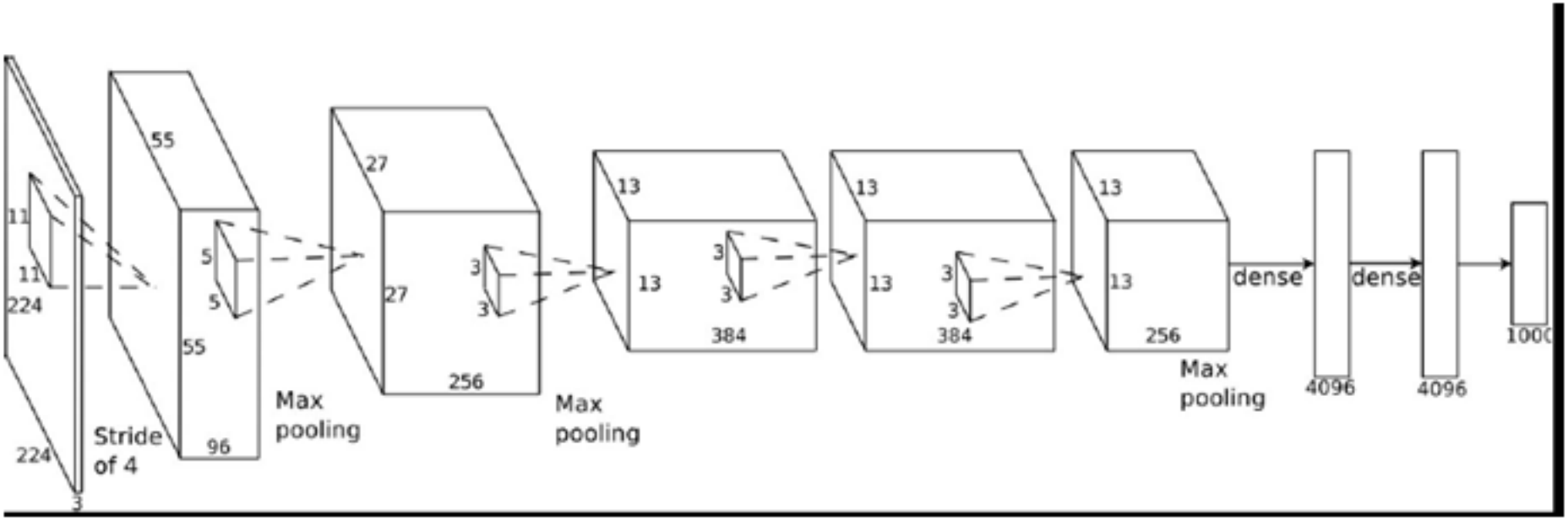
Свёрточные сети (1998)



- «Свёрточный слой» реализует операцию свёртки
  - Каждой области одного размера на предыдущем слое соответствует свой нейрон
  - Вектор весов у всех нейронов одного слоя одинаковый
- Это резко снижает число признаков

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

SuperVision (2012)



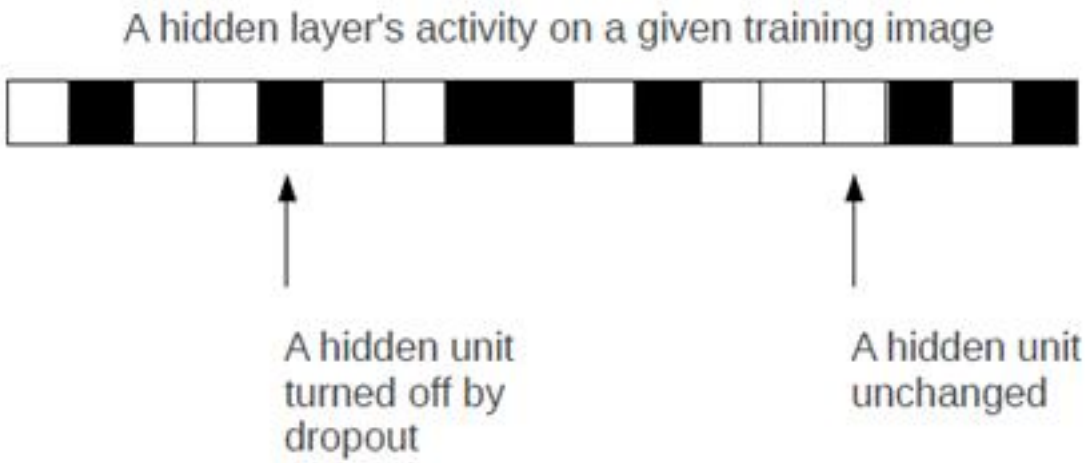
- 650,000 neurons
- 60,000,000 parameters
- 630,000,000 connections
- 1 машина, 2 GPU по 2Gb, 5GB Ram, 27Gb HDD, 1 неделя на обучение на LSVRC

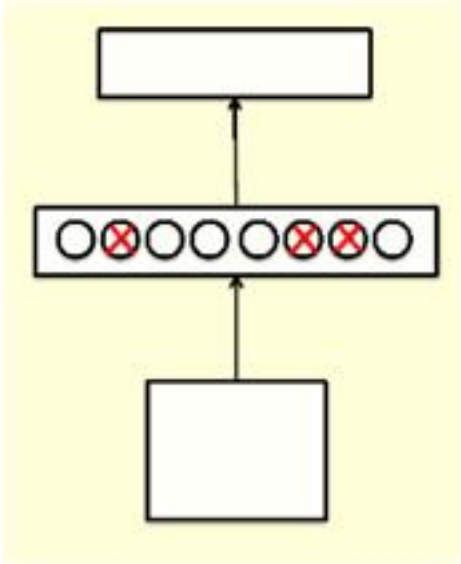
Krizhevsky A., Sutskever I., Hinton, G. E. (2012) ImageNet Classification with Deep Convolutional Neural Networks // NIPS 2012: Neural Information Processing Systems. Lake Tahoe, Nevada.

Размножение данных

- Борьба с переобучением
- Из 256x256 случайно выбираем фрагменты 224x224 и их отражения

Dropout

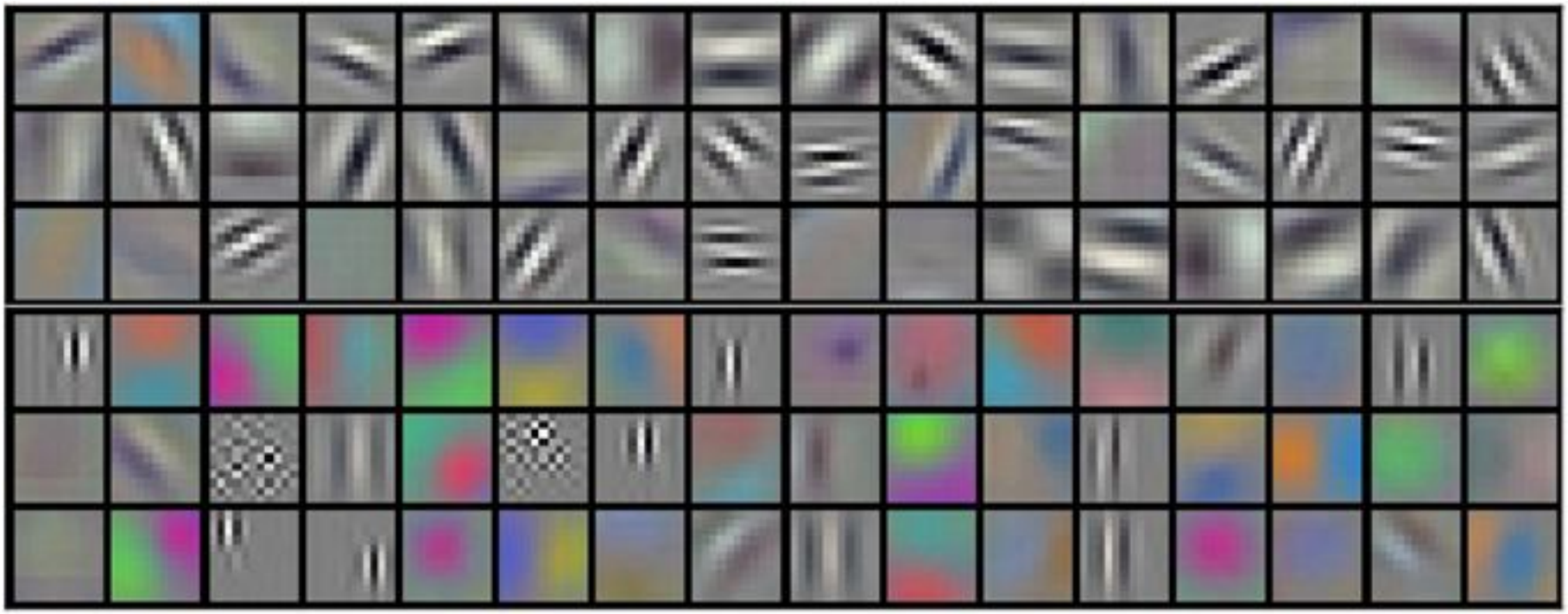




- Отключаем половину нейронов в каждом слое
- Получаем случайную выборку из множества сетей
- Во время тестирования используем «среднюю» сеть с уполовиненными весами

Nitish Srivastava Improving Neural Networks with Dropout. Master Thesis, 2013

Фильтры первого уровня



Примеры работы

Примеры работы

Прокрастинация или послойное обучение

SexNet

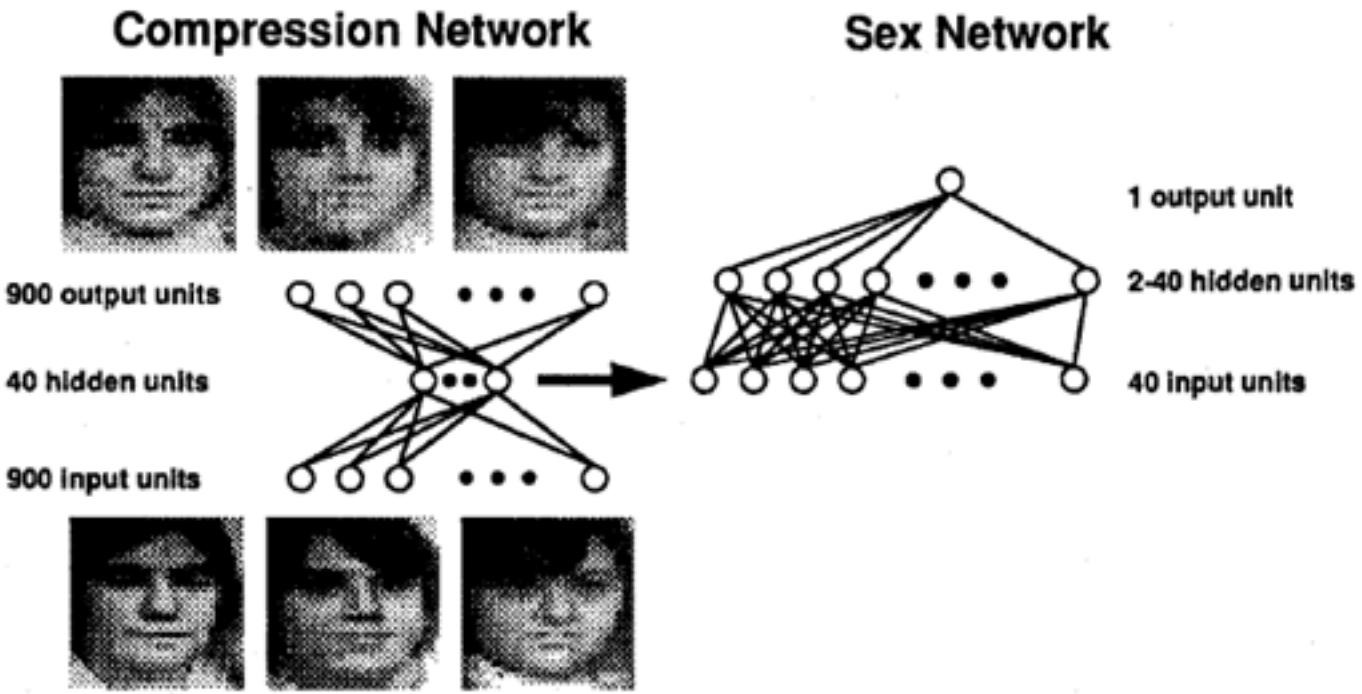


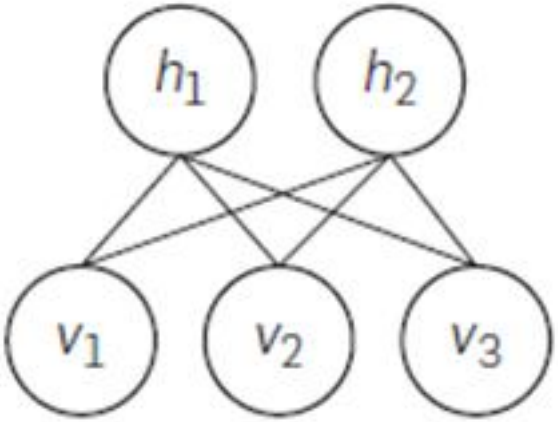
Figure 2: Two-stage network for discrimating sex from faces. The compression network encodes the normalized faces into 40 hidden units, which are then used as inputs to a sex network. The 30x30 input image has 256 gray levels per pixel. The output of the sex network is 1 for male and 0 for female.

Сеть сжатия – подбираем веса скрытого слоя таким образом, чтобы можно было восстановить исходный сигнал

Touretzky, Lippman. Sexnet: a neural network idendifies sex from human faces. 1991

RBM

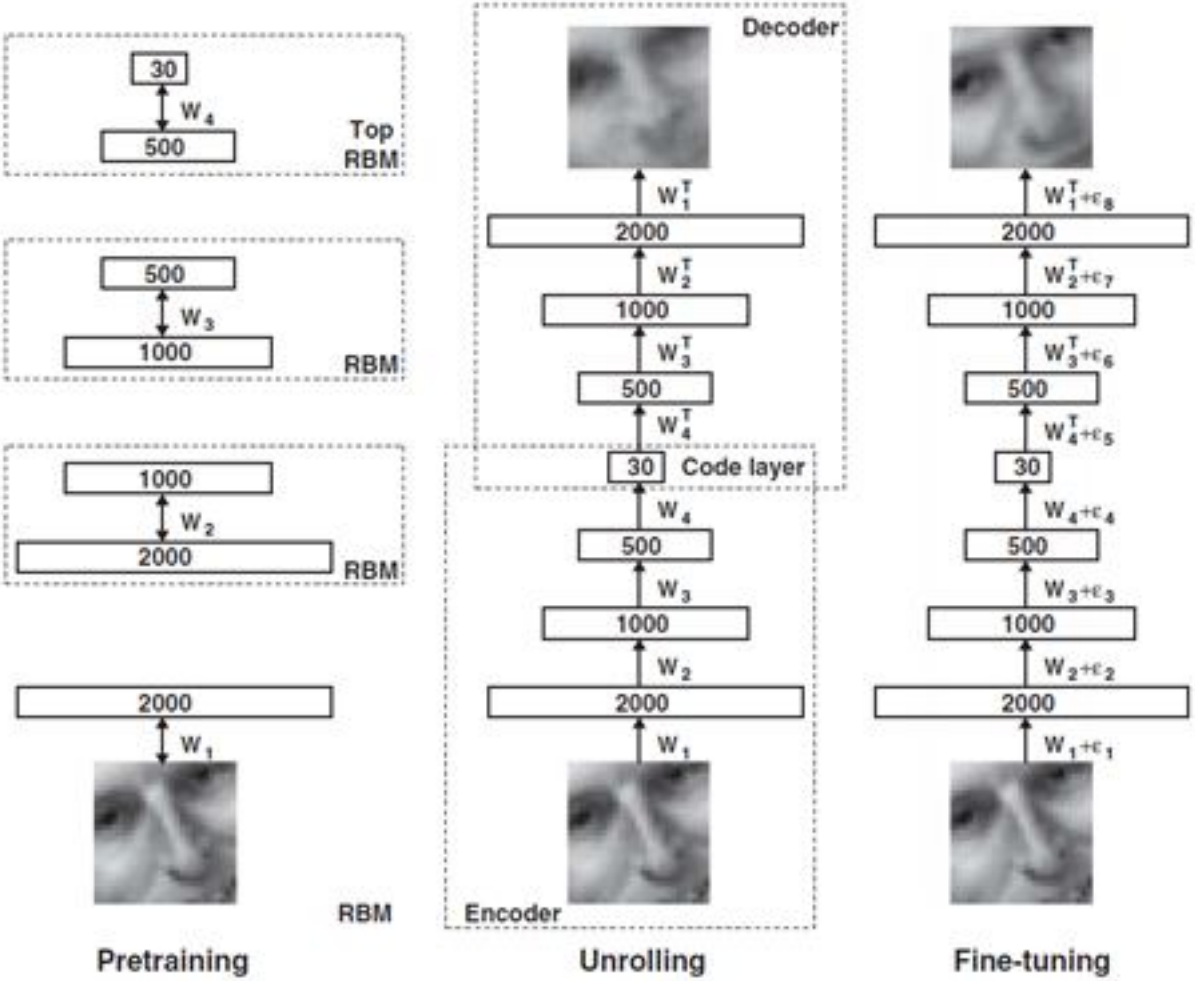
- Restricted Boltzmann Machine
- Марковская сеть на двудольном графе



- Фактически, «сжатие» данных

H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng. Unsupervised learning of hierarchical representations with convolutional deep belief networks. Communications of the ACM, 54(10):95–103, 2011.

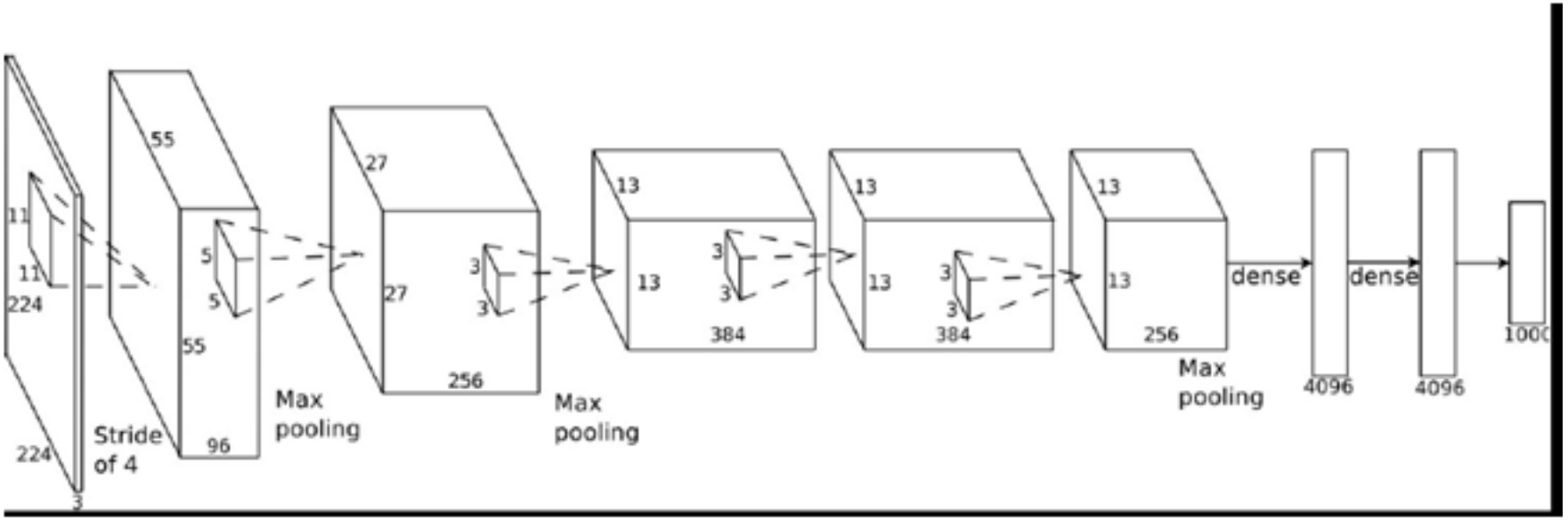
Deep Belief Networks



- Объединение нескольких RBM в многослойную структуру
- Каждый слой пре-обучается с помощью алгоритма autoencoder
- Затем вся структура уточняется с помощью градиентного спуска

Hinton, G.E., Salakhutdinov, R. Reducing the dimensionality of data with neural networks. *Science 313*, 5786 (2006), 504–507.

Разреженные фильтры



- Обучение фильтров через К-средних
- Специальное разреженное представление

Свёрточный разреженный базис

$$z^* = \arg\min_z \|x - \mathcal{D}z\|_2^2 + \lambda s(z)$$

$$z^*, \mathcal{D}^* = \arg\min_{z,\mathcal{D}} \|x - \mathcal{D}z\|_2^2 + \lambda \|z\|_1$$

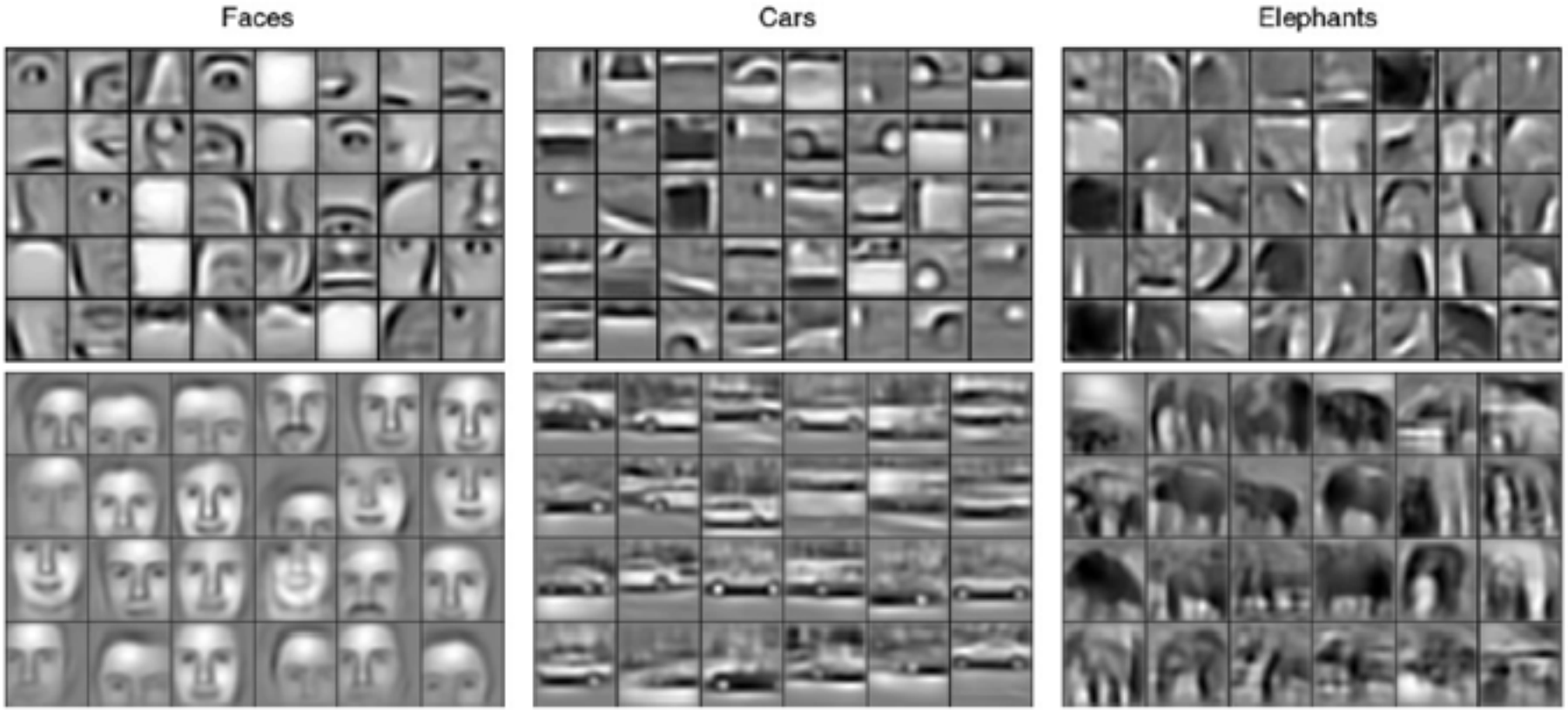
$$f(x; g, k, b) = \tilde{z} = \{\tilde{z}_j\}_{j=1..n}$$
$$\tilde{z}_j = g_j \times \tanh(x \otimes k_j + b_j)$$

$$\mathbb{E}_{CPSD} = \mathbb{E}_{ConvSC} + \beta \mathbb{E}_{Pred}$$
$$\mathbb{E}_{ConvSC} = \left\| x - \sum_j \mathcal{D}_j \otimes z_j \right\|_2^2 + \lambda \|z\|_1$$
$$\mathbb{E}_{Pred} = \|z^* - f(x; g, k, b)\|_2^2$$

- Обучение разреженного базиса как набора линейных фильтров

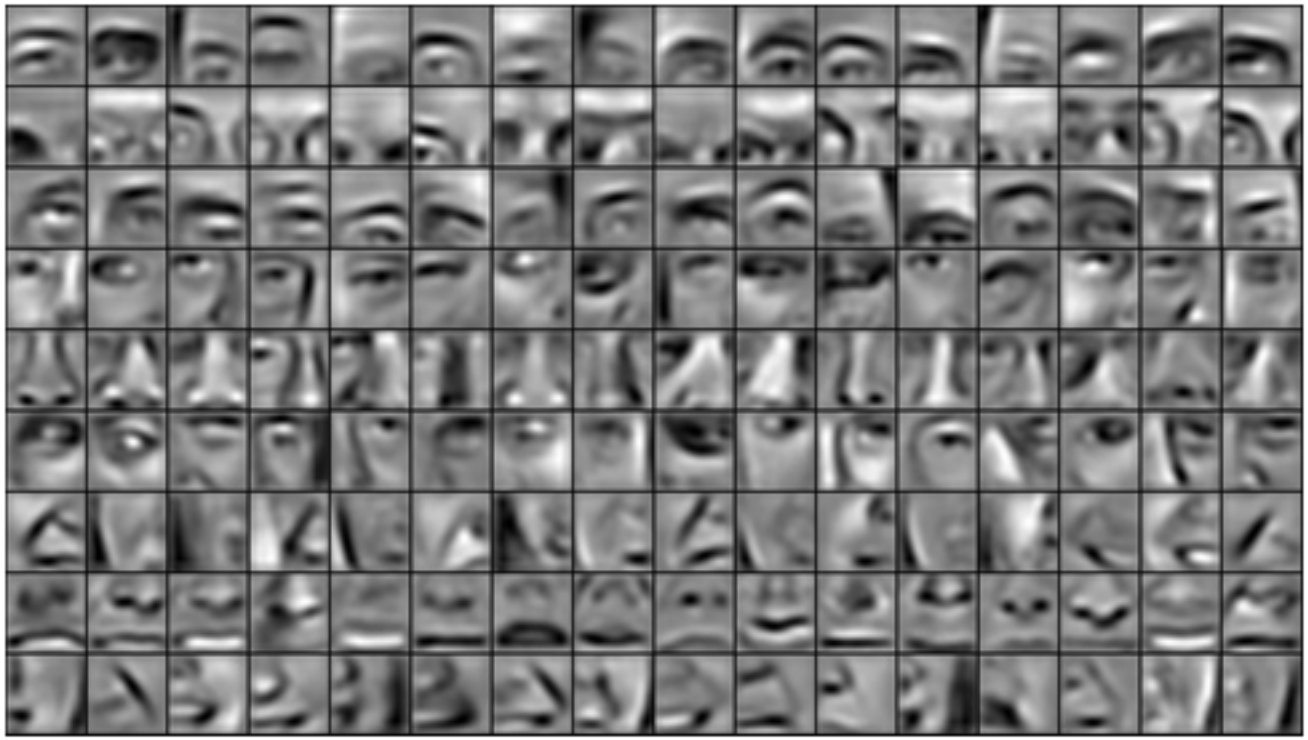
Pierre Sermanet, Koray Kavukcuoglu, Soumith Chintala,Yann LeCun Pedestrian Detection with Unsupervised Multi-Stage Feature Learning, CVPR 2013

Примеры



- Фильтры, который были найдены при обучении на определенных классах изображений

Local Convolutional RBM



- Разделим изображение на перекрывающиеся фрагменты
- Веса теперь будут одинаковы только в пределах фрагмента
- Сможем учесть локальные особенности каждого региона

Huang et.al. Learning hierarchical representations for face verification with convolutional deep belief networks, CVPR 2012

Резюме

- Нейросети заметно изменились, но топ-результаты можно получить и используя старый подход (свёрточные)
- Ключевой момент – объём данных и вычислительных ресурсов для обучения
- Важный вопрос, насколько их можно использовать сейчас для обнаружения объектов с достаточной скоростью
  - Активно изучают создание специальный процессоров для вычисления нейросетей

Thank you for evaluating AnyBizSoft PDF Splitter.

A watermark is added at the end of each output PDF file.

To remove the watermark, you need to purchase the software from

http://www.anypdftools.com/buy/buy-pdf-splitter.html