

Школа Анализа
Данных Яндекса

Курс «Анализ изображений и видео, ч.2.»

Лекция №3
«Сегментация изображений»

Антон Конушин

Заведующий лабораторией компьютерной графики и мультимедиа
ВМК МГУ

3 марта 2017 года

Задача сегментации



Разделение изображения на фрагменты, группы пикселов по определённому критерию



Извлечение объекта



Выделение конкретного произвольного объекта, указанного пользователем или по другому заданного



Сегментация без учителя



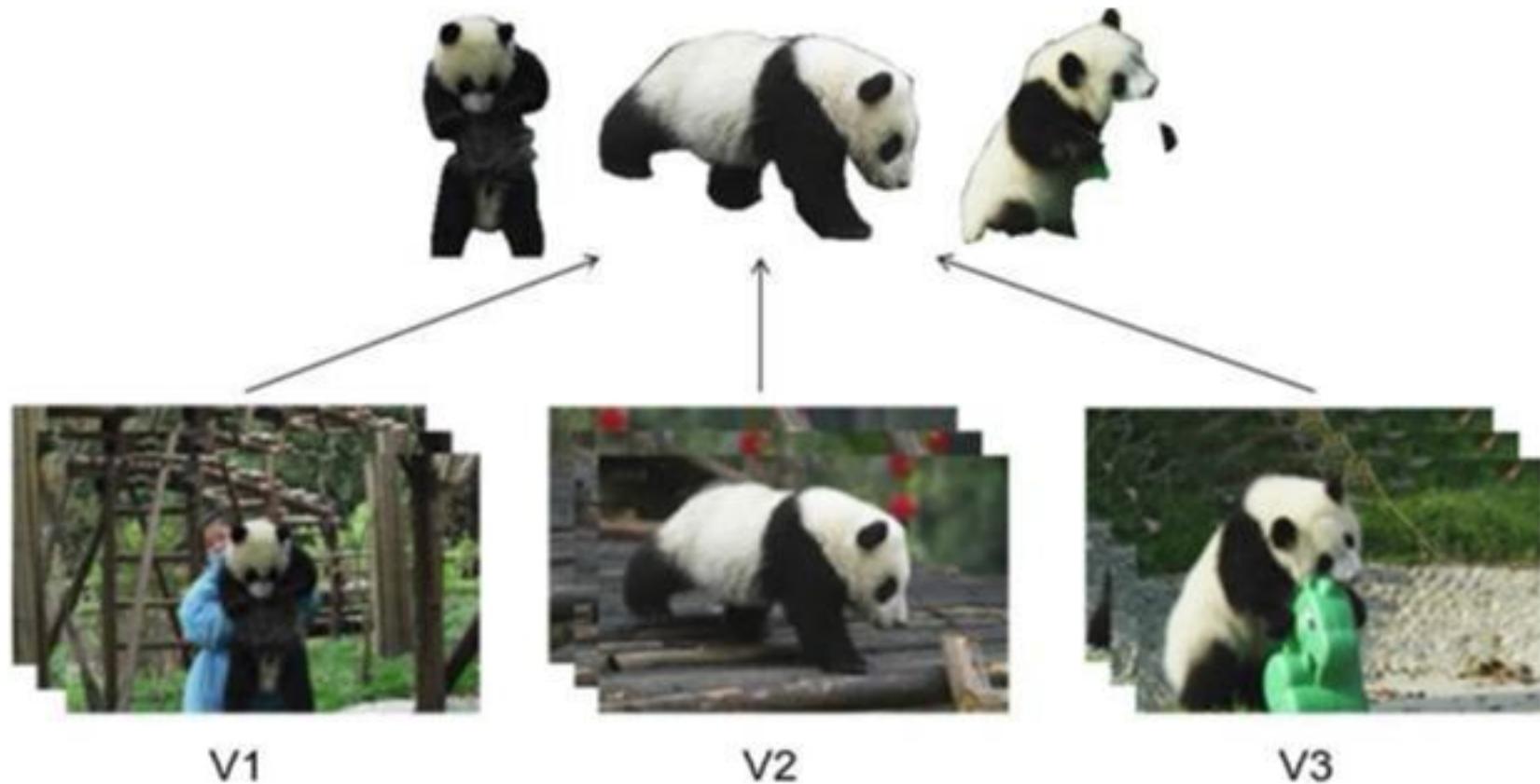
Разделение изображения на регионы, однородные по своим визуальным характеристикам, и отличающихся от соседних регионов



Ко-сегментация



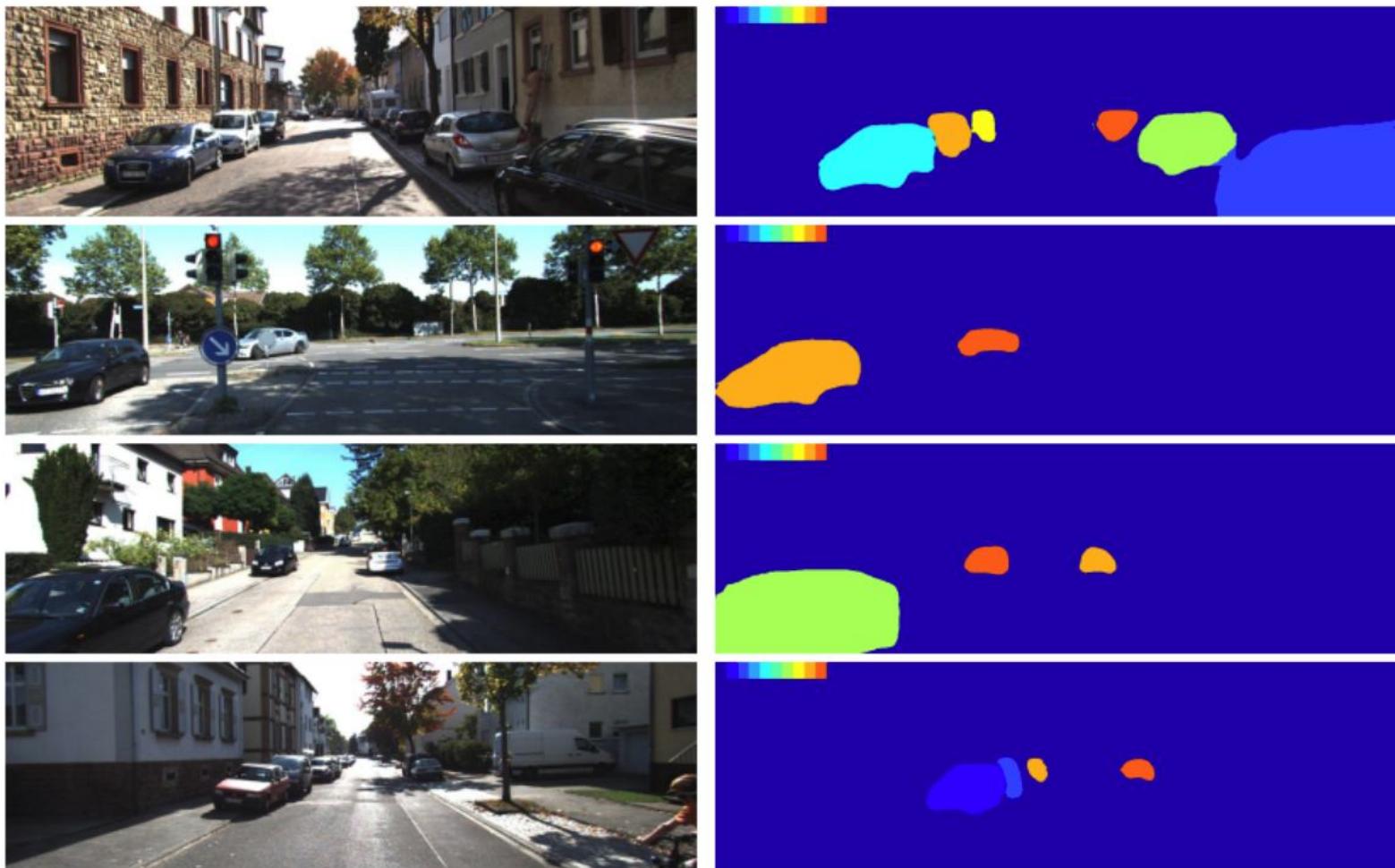
Выделение объекта, который присутствует на всех изображениях из выборки



Instance segmentation



Выделение всех отдельных экземпляров объектов определённого класса. Каждый объект помечен своей меткой.



Семантическая сегментация

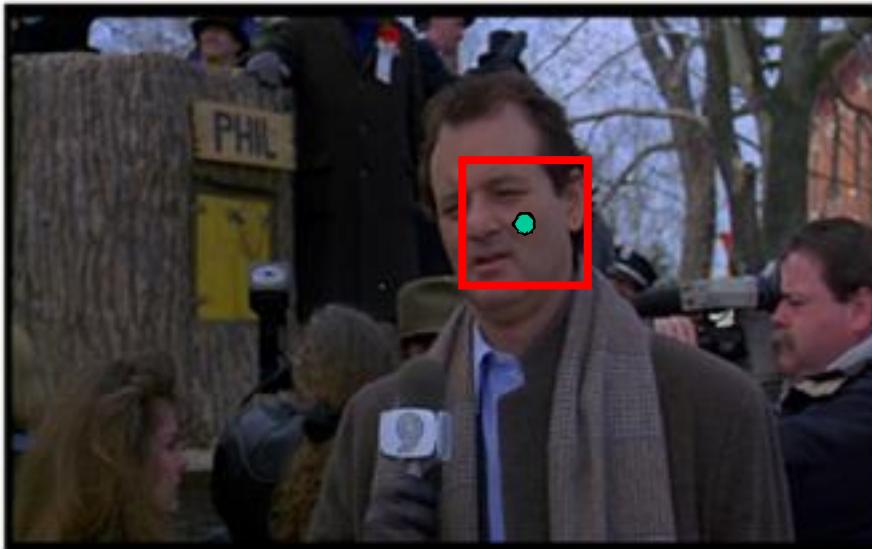


Попиксельная разметка изображения, где каждая метка соответствует определенному объекту



- Разные пиксели одного объекта существенно отличаются друг от друга по признакам (яркости, цвету, текстуре окрестности)
- Единственное, что у них общее – это «семантика»
- Поэтому задача сегментации объекта тесно связана с задачей распознавания

Простейший способ решения



Независимая попиксельная (локальная) классификация:

- Применяем классификатор к каждому пикслю
- Постобработка результата

Точность семантической сегментации в первую очередь зависит от точности локальных классификаторов.

Но есть и ряд других влияющих факторов.



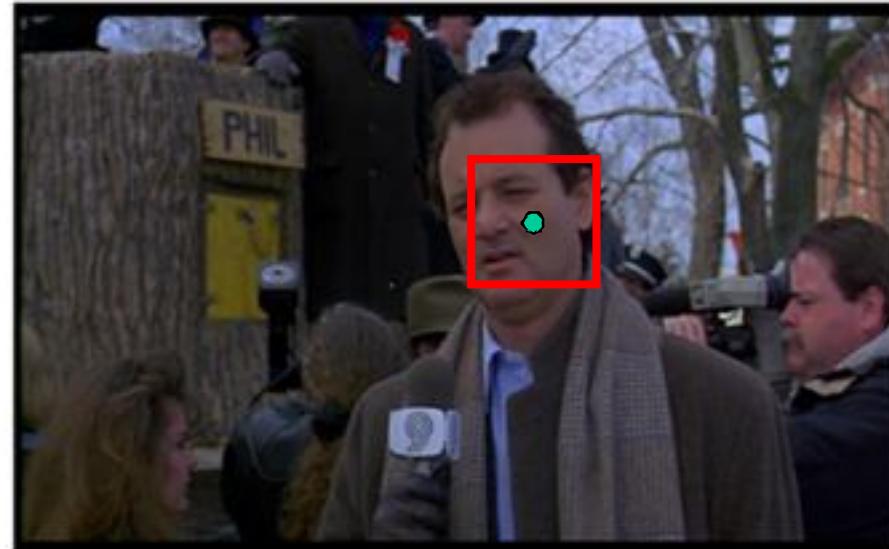
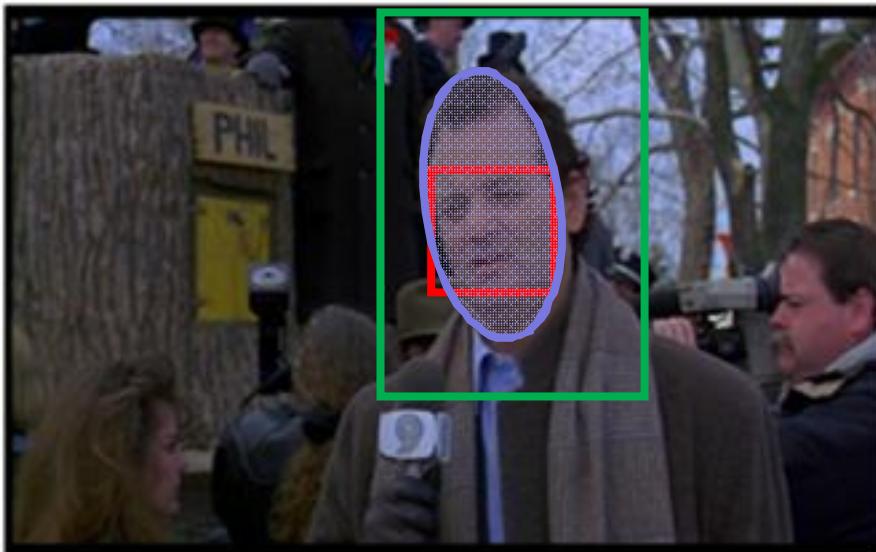
Факторы

Можно выделить набор факторов, влияющих на семантическую сегментацию:

- Пространственная поддержка
- Локальный контекст
- Глобальный контекст

Рассмотрим эти факторы. Затем посмотрим некоторые подходы к решению визуальной сегментации и семантической сегментации.

Пространственная поддержка

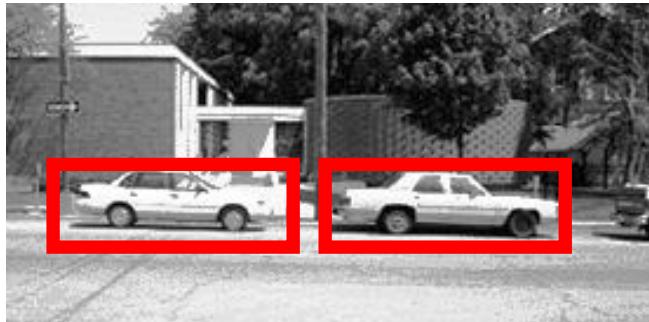


- Область, по которой вычисляются признаки при распознавании - пространственная поддержка (Spatial Support)
- Принимать решение мы можем для области целиком или для точки или фрагмента внутри области
- Область можем брать заранее определённой формы или как-то определять форму по данным
- Насколько пространственная поддержка важна?

Распознавание и сегментация



Машины

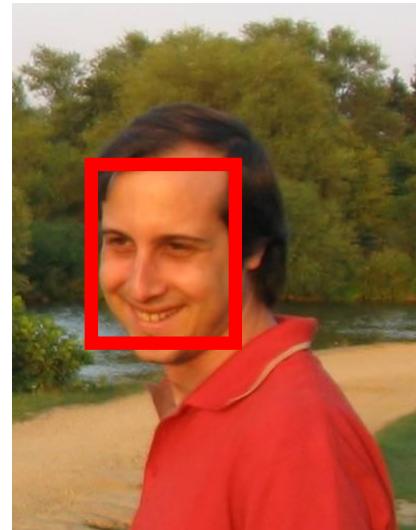


Schneiderman & Kanade '00

Мы рассматривали
методы выделения
объектов на основе
скользящего окна

Скользящее окно – это
вид сегментации!

лица



Viola & Jones '04
Schneiderman & Kanade '00

пешеходы



Dalal & Triggs '05
Ferrari et al '07

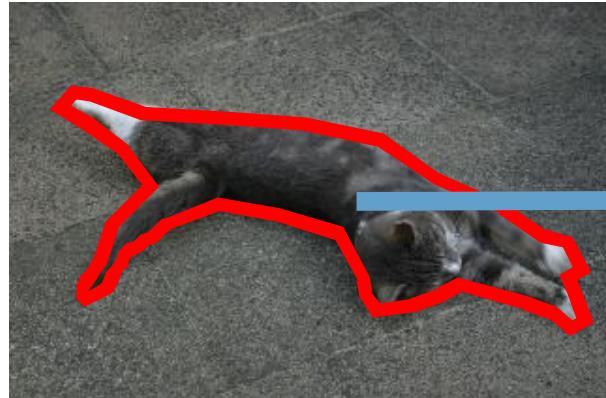
Slide by Alexey Efros

Ограничения скользящего окна



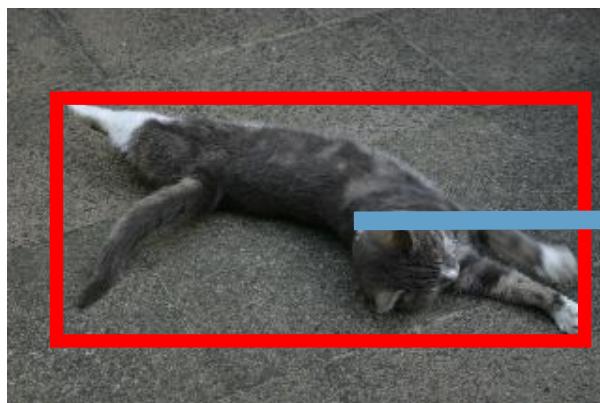
Форма многих объектов сильно далека от прямоугольника. В рассматриваемую область попадает много пикселей других объектов

Насколько важна поддержка?



Эталонная разметка

Classify



Ограничивающий
прямоугольник

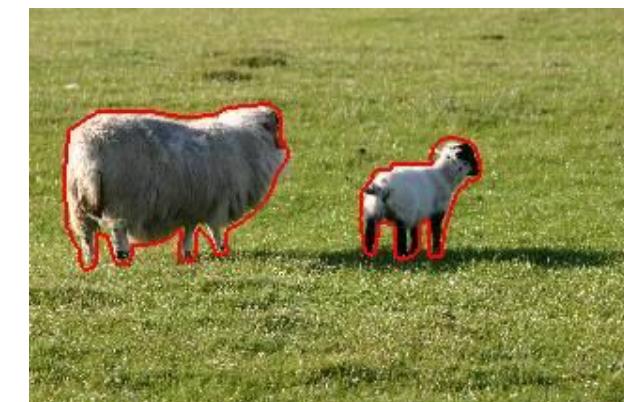
Проведём эксперимент
и сравним подходы

Classify

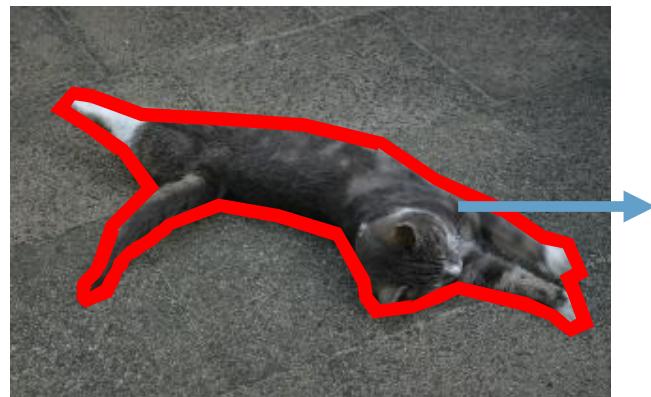
Насколько важна поддержка?



База MSRC: 591 картинок 23 класса объектов +
Попиксельные маски объектов



Насколько важна поддержка?



Признаки

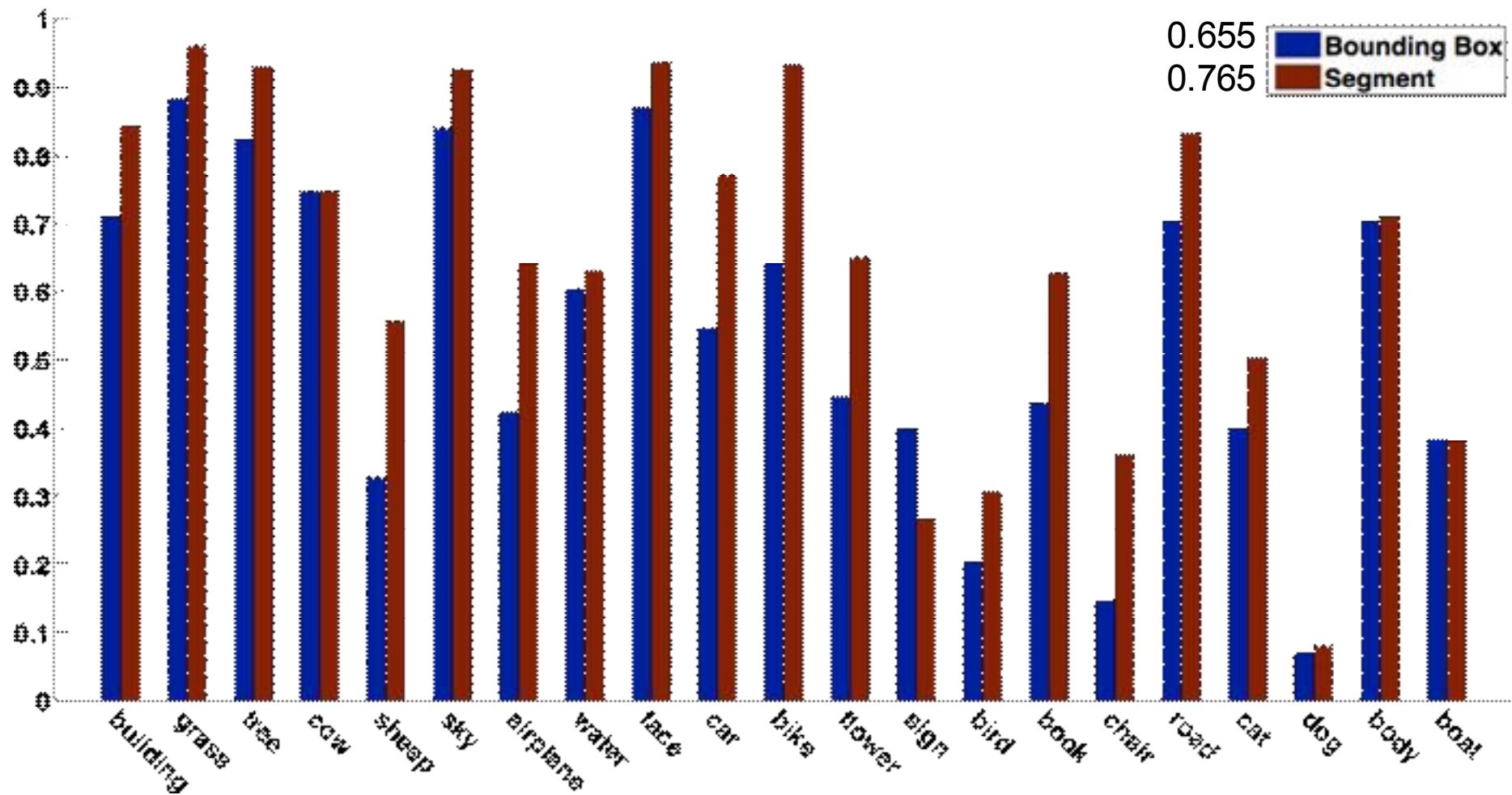
*

Feature Descriptions	Num
Color	16
C1. RGB values: mean	3
C2. HSV values: C1 in HSV space	3
C3. Hue: histogram (5 bins) and entropy	6
C4. Saturation: histogram (3 bins) and entropy	4
Texture	15
T1. DOOG filters: mean abs response of 12 filters	12
T2. DOOG stats: mean of variables in T1	1
T3. DOOG stats: argmax of variables in T1	1
T4. DOOG stats: (max - median) of variables in T1	1
Location and Shape	12
L1. Location: normalized x and y, mean	2
L2. Location: norm. x and y, 10 th and 90 th pctl	4
L3. Location: norm. y wrt horizon, 10 th , 90 th pctl	2
L4. Shape: number of superpixels in region	1
L5. Shape: number of sides of convex hull	1
L6. Shape: <i>num pixels/area(convex hull)</i>	1
L7. Shape: whether the region is contiguous $\in \{0, 1\}$	1
3D Geometry	35
G1. Long Lines: total number in region	1
G2. Long Lines: % of nearly parallel pairs of lines	1
G3. Line Intscn: hist. over 12 orientations, entropy	13
G4. Line Intscn: % right of center	1
G5. Line Intscn: % above center	1
G6. Line Intscn: % far from center at 8 orientations	8
G7. Line Intscn: % very far from center at 8 orient.	8
G8. Texture gradient: x and y "edginess" (T2) center	2

Классификатор

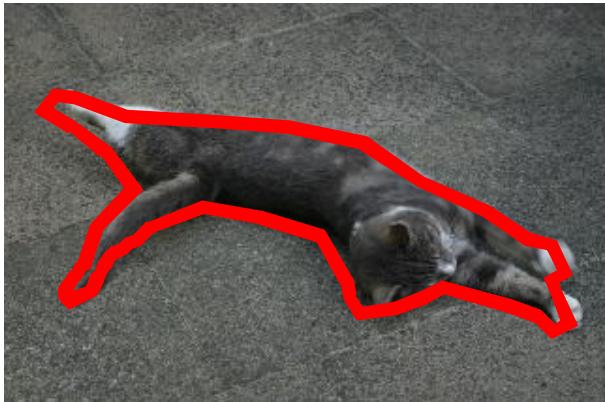
Boosted Decision Tree*

Насколько важна поддержка?



Для многих объектов выбор правильной области для пространственной поддержки повышает качество распознавания

Сегментация и распознавание



- Распознавать объект проще, если правильно выделена область
- Область объекта можно определить, если объект распознан
- Проблема «курицы и яйца»
- Возможно, это не так важно для разреженных высокоуровневых представлений, который строятся в нейросетевых моделях

Локальный контекст



- Рассмотрим несколько пар точек на изображении
- Что мы можем сказать про их принадлежность одним и тем же объектам и почему?

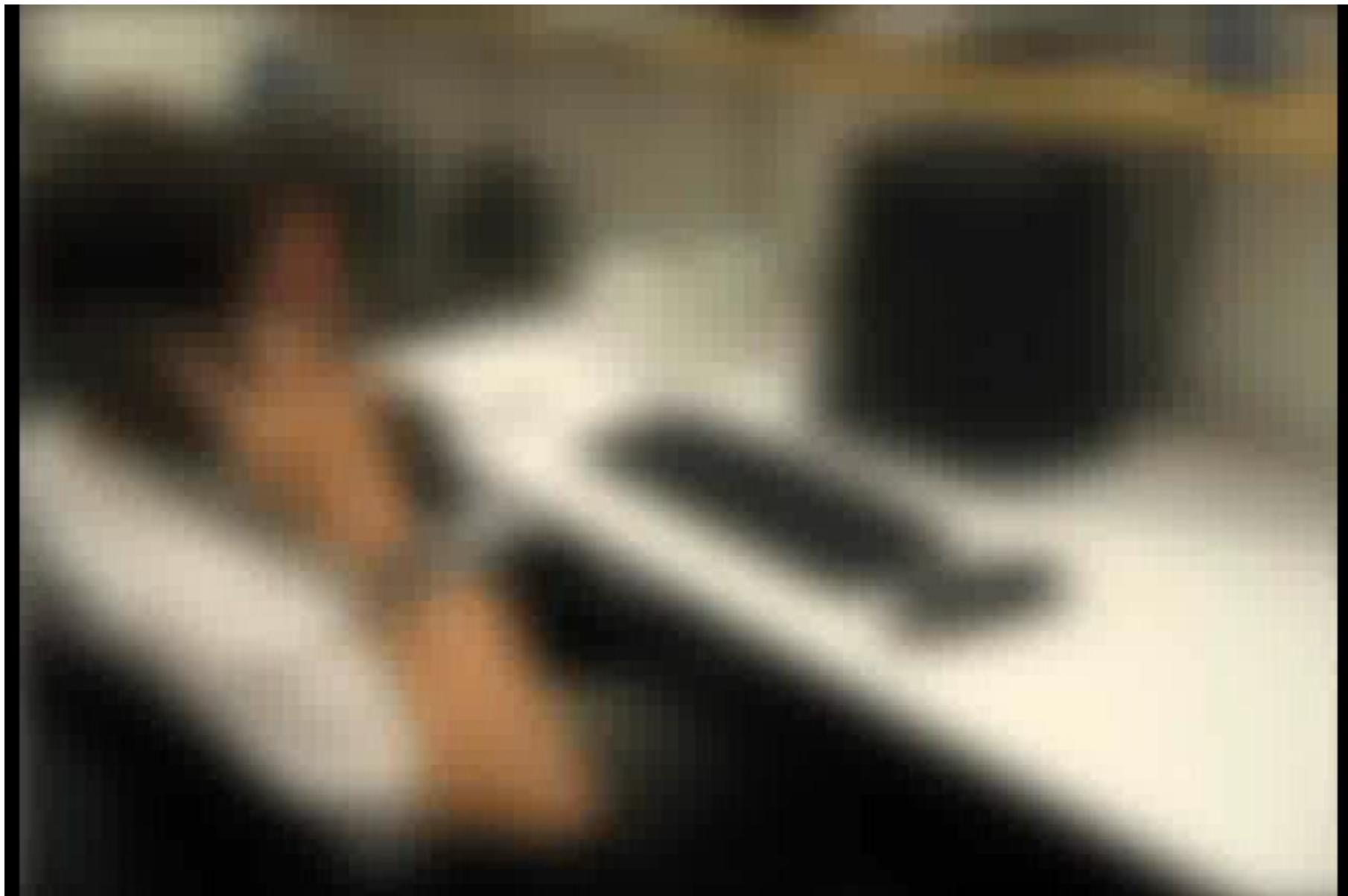
Локальный контекст



- Если точки одинакового цвета – выше вероятность принадлежности одному объекту
- Если между точками есть граница – ниже вероятность принадлежности одному объекту
- Некоторые границы разделяют разные объекты, некоторые соответствуют текстуре объекта



Тест





Тест



Что изображено на картинках?



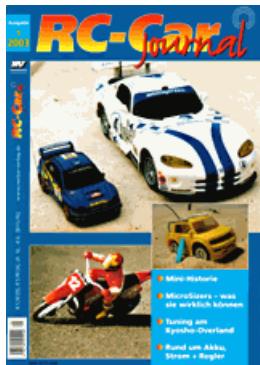
Что изображено на картинках?



Влияние глобального контекста



- Изменяет интерпретацию объекта (или функцию)



- По контексту определяются «внештатные» ситуации



Важность контекста



Figure 9.1.2 Objects with similar structure but different functions. Mailboxes afford letter mailing, whereas trash cans do not, even though they have many similar physical features, such as size, location, and presence of an opening large enough to insert letters and medium-sized packages.



Объекты одной формы и вида могут иметь разное назначение



Визуальная сегментация

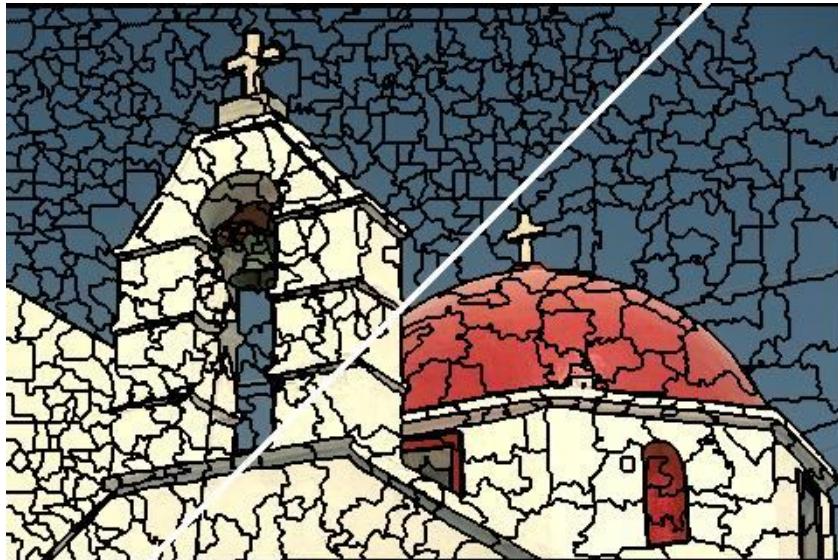
Требования к сегментации?



Каким условиям должны удовлетворять сегменты?

- Границы сегментов должны соответствовать границам объектов
- Сегмент должен целиком содержаться внутри объекта
- Небольшие объекты не должны быть частью сегмента, а описываться своим сегментом
- Сегмент должен быть однородным по визуальным характеристикам
- Сегменты должны быть достаточно большими, чтобы быть «информационными»
- Компактные, примерно одного размера
- Равномерно распределенные по изображению
- Алгоритм должен работать быстро

Суперпиксели



- Сегменты, «удовлетворяющие» вышеуказанным требованиям называют «суперпиксели»
- «Суперпиксельная сегментация»
- Ещё называют «пересегментацией» (oversegmentation)

Как оценить качество сегментации?



- Точность сегментации
 - Границы сегментов должны соответствовать границам объектов (численно)
 - Сегмент должен целиком содержаться внутри объекта (численно)
 - Компактность сегментов (визуально)
- Удобство использования
 - Возможность регулировать число сегментов
 - Скорость работы
 - Требования по памяти

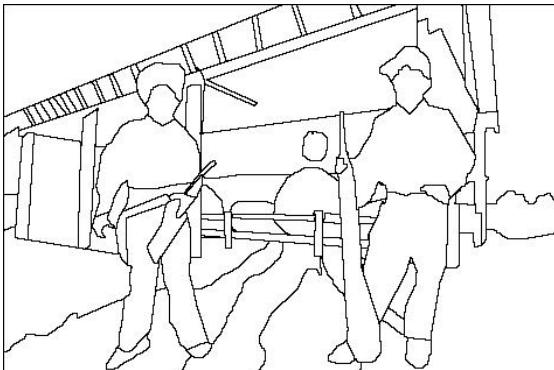
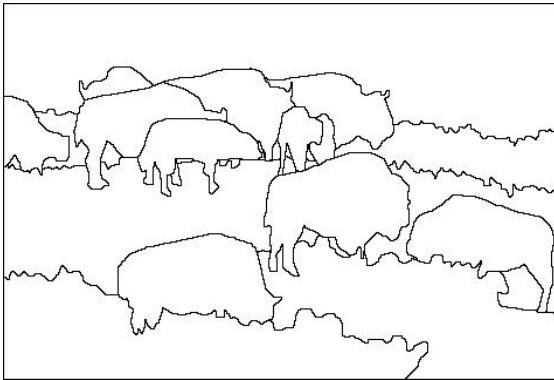
Точность границ



изображение



ручная разметка



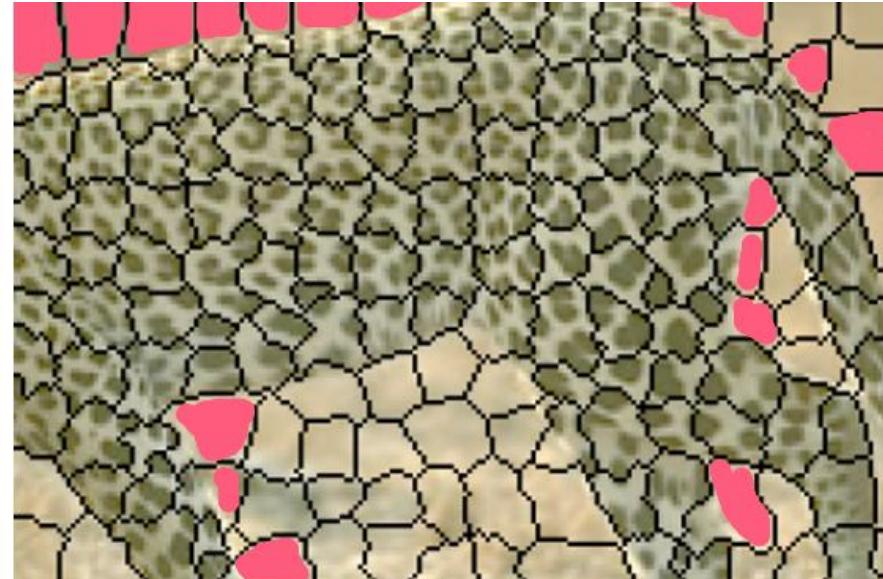
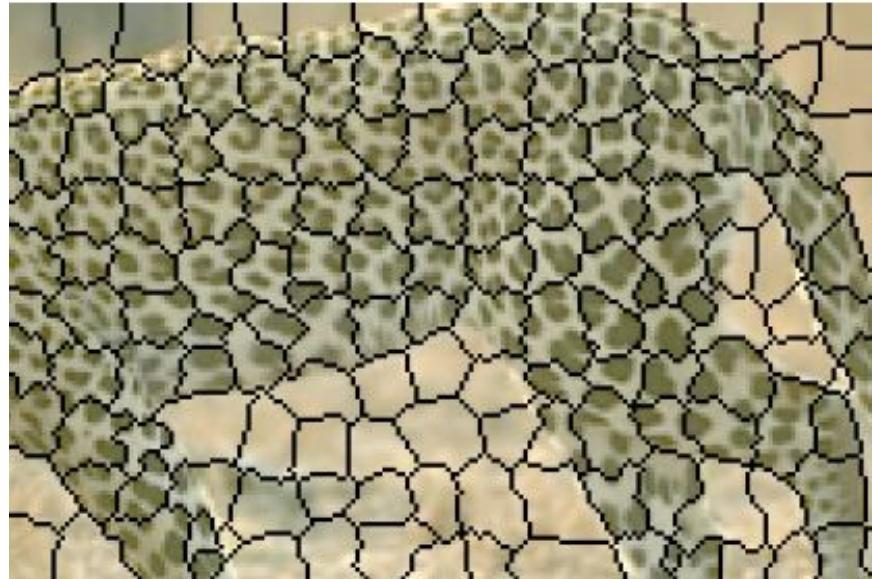
Сравнение границ сегментов с эталонными границами, выделенными людьми (recall)

Нужно вводить порог на близость – из-за неточности эталонной разметки

Berkeley segmentation database:

<http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/>

Точность сегментации



- Доля сегментов, которые выходят за границы эталонной разметки, но пересекаются с ней
- «undersegmentation»

Подходы к сегментации



- Эвристические методы
 - Разрастание регионов (Region growing)
 - Разделения и слияния регионов (Split & Merge)
- Методы на графах
 - Метод Felzenszwalb & Huttenlocher
- Энергетические методы
 - Snakes
 - Методы уровня - ТурбоПиксели (TurboPixels)
- Кластеризация
 - Сдвиг среднего (Mean shift)
 - Быстрый сдвиг (QuickShift)
 - SLIC



Мы рассмотрим

- Сегментация на основе кластеризации
 - Сдвиг среднего (Mean shift) и быстрый сдвиг (QuickShift)
 - SLIC
- Описание текстуры
 - Текстоны – признаки текстуры
 - Рb-детектор краёв

Пересегментация через кластеризацию



- Представим множество пикселов как выборку (множество вектор-признаков)
- Применим какой-нибудь метод кластеризации к данным в пространстве признаков
- Каждый кластер будет соответствовать одному сегменту
- Как обеспечить «локальность» кластеров в изображении?
 - (x, y) можно включить в вектор-признак
 - Жесткие ограничения на расстояния между пикселями



К-средних



Исходное изображение



Кластеры по яркости



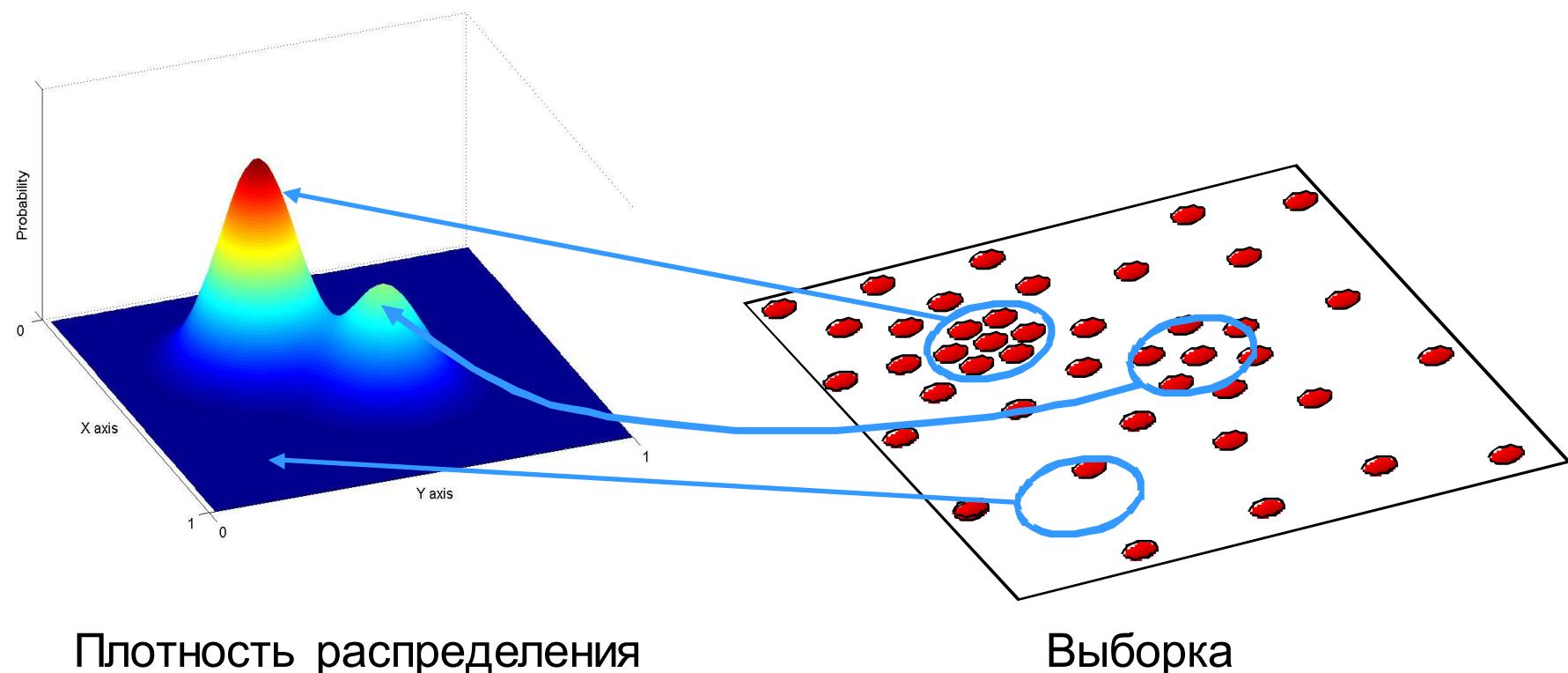
Кластеры по цвету

Можно воспользоваться обычным методом К-средних. Какие у него недостатки?

Сдвиг среднего (mean shift)

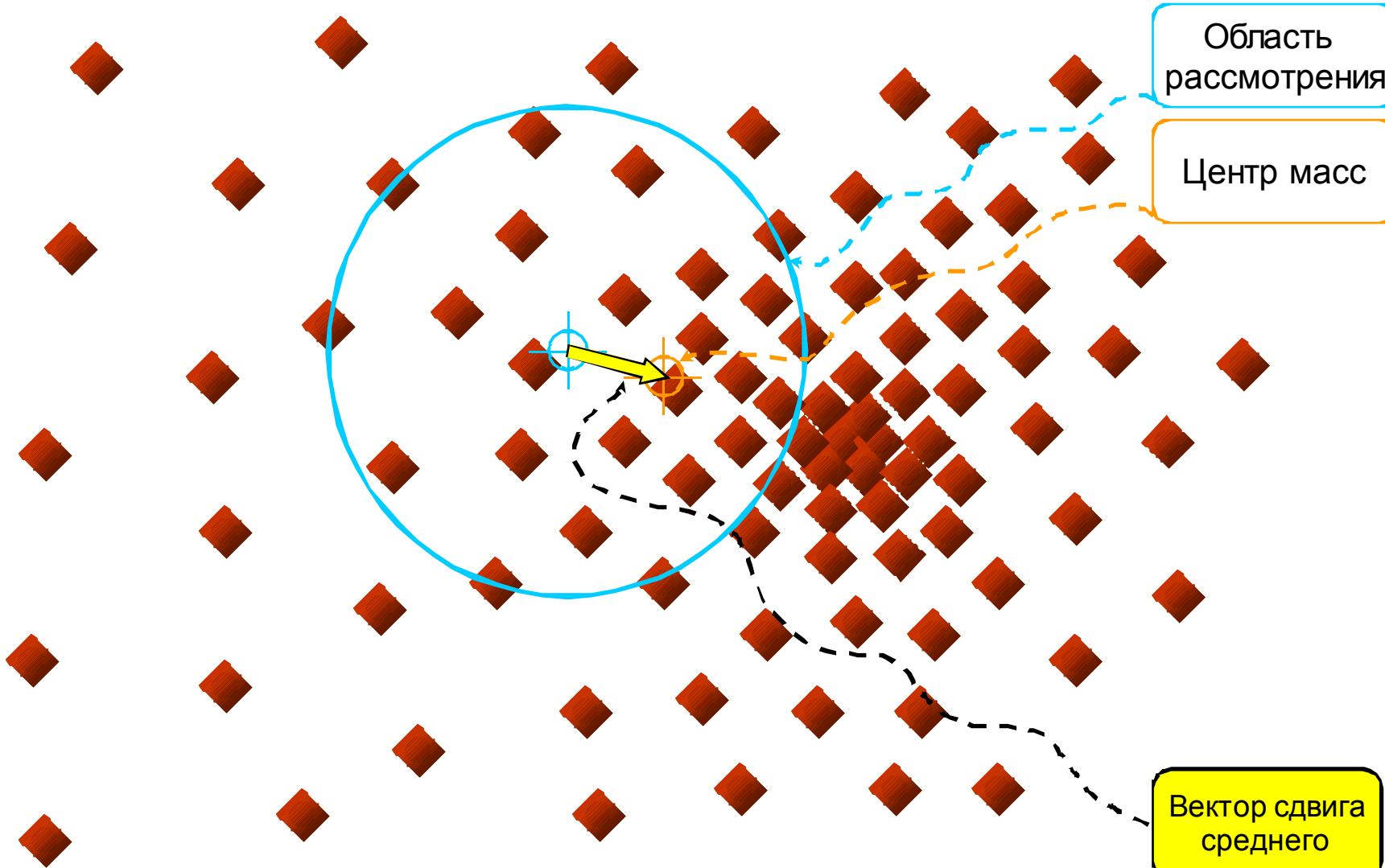


- Кластеры – плотно сгруппированные в пространстве признаков группы пикселей
- Идея - центры кластеров должны соответствовать локальным максимумам плотности распределения элементов данных



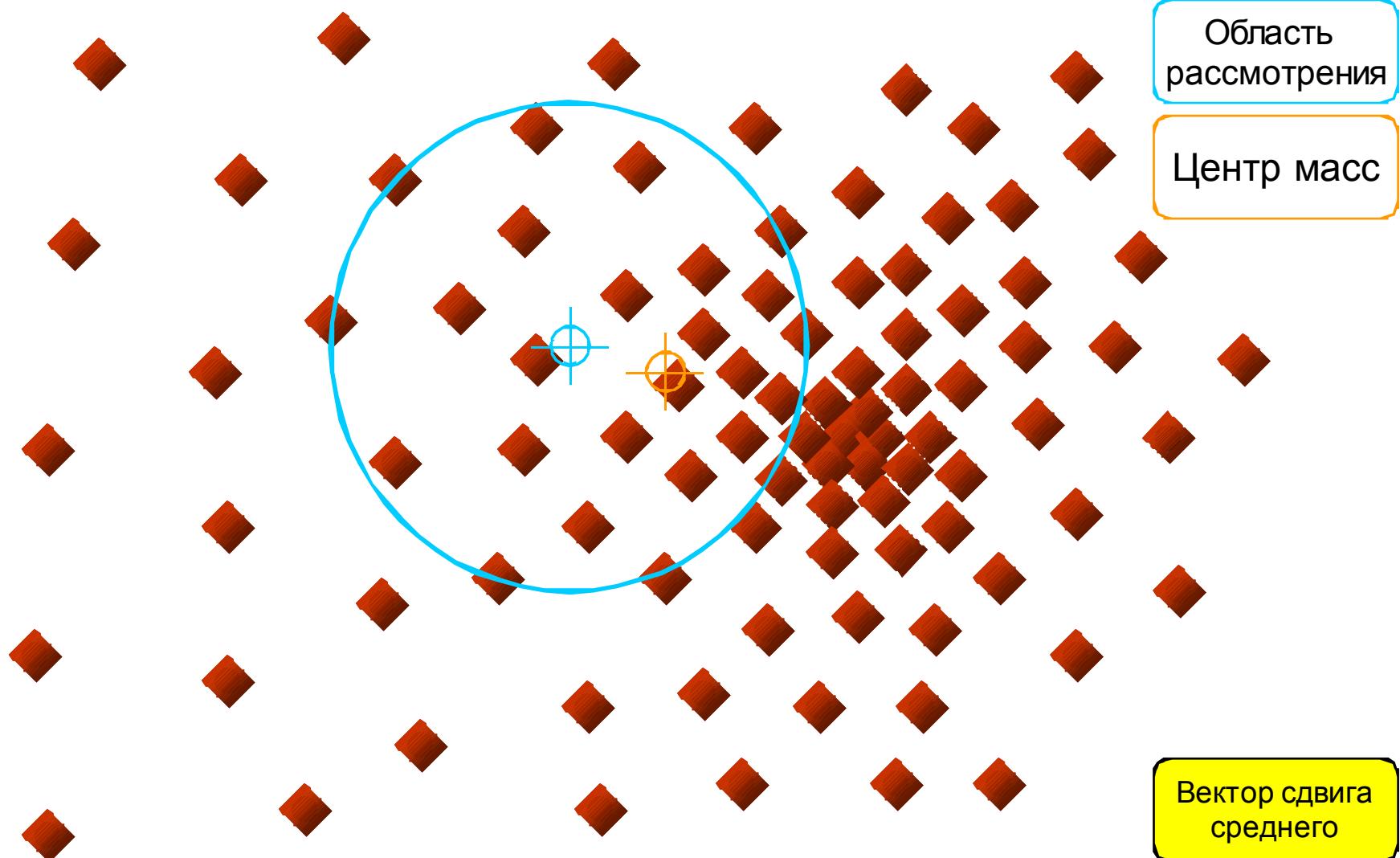
Comaniciu, Dorin; Peter Meer "Mean Shift: A Robust Approach Toward Feature Space Analysis". *IEEE Transactions on Pattern Analysis and Machine Intelligence* (IEEE), **24** (5): 603–619. 2002 (>6500 цитирований)

Метод сдвига среднего: общая идея



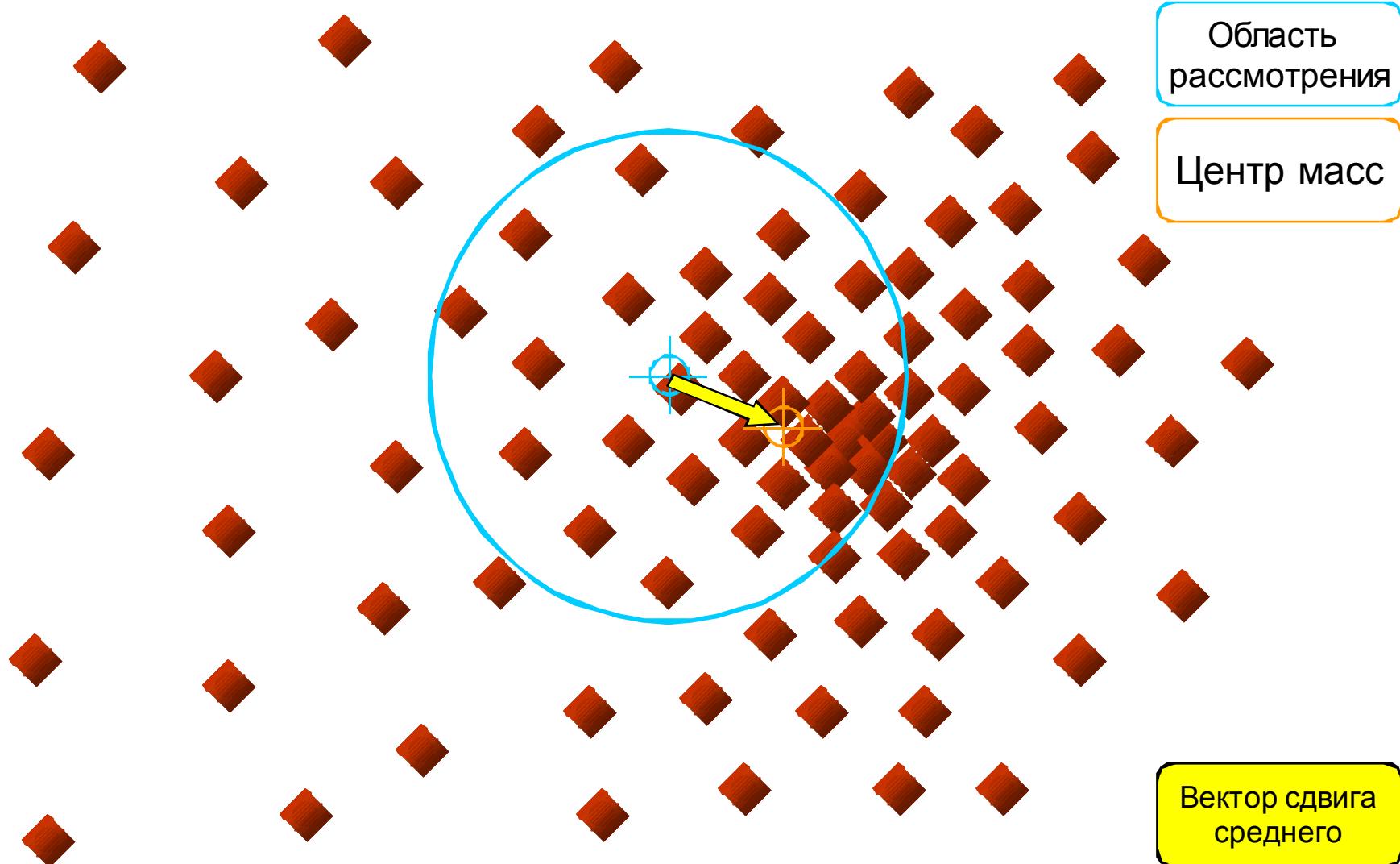
Можно показать, что сдвиг средней радиуса традиционно эквивалентен
дой сдвигу в точку из максимальной среднестатистической окрестности

Метод сдвига среднего: общая идея

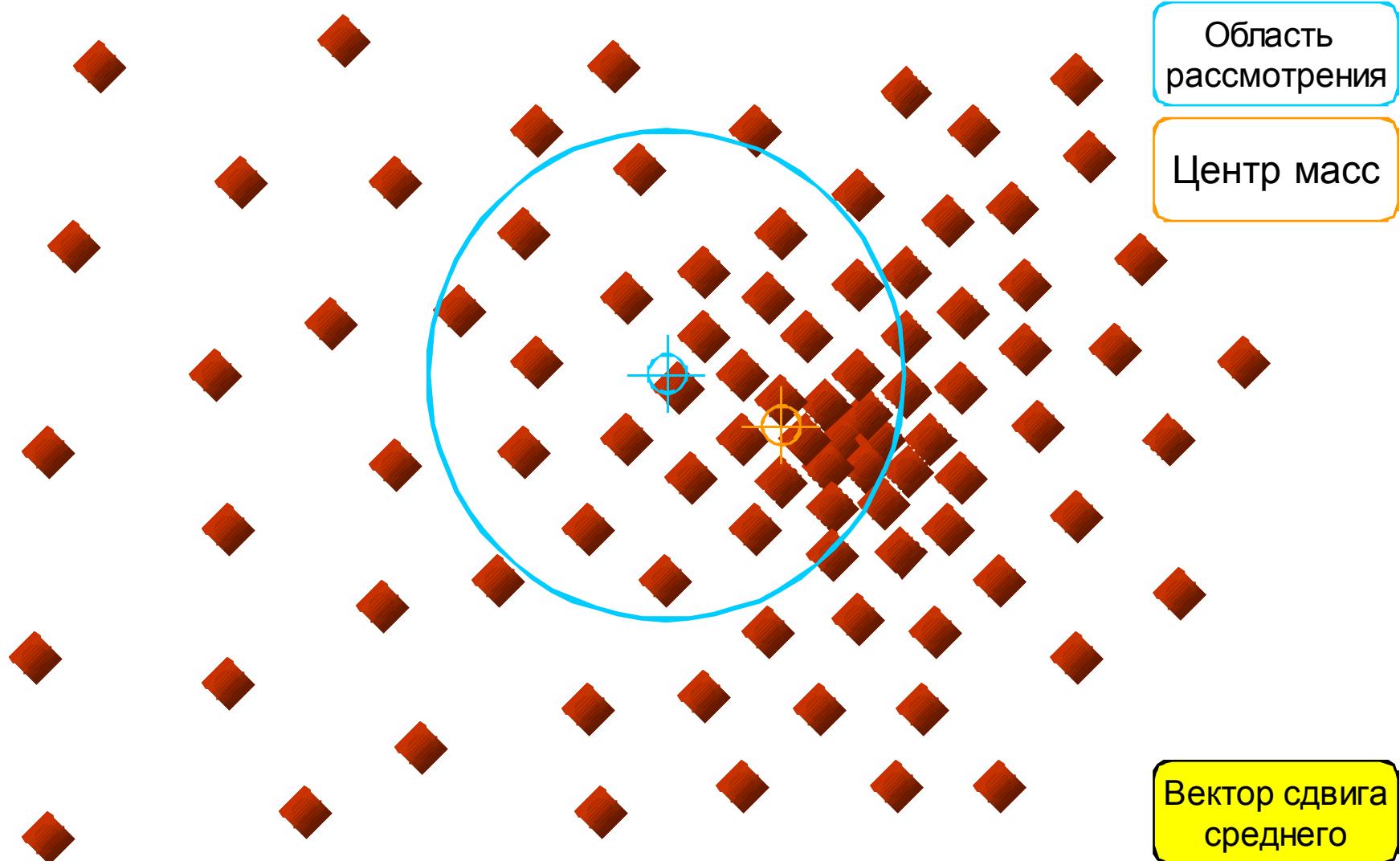


Можно показать, что сдвиг по градиенту плотности эквивалентен сдвигу в точку взвешенного среднего по окрестности

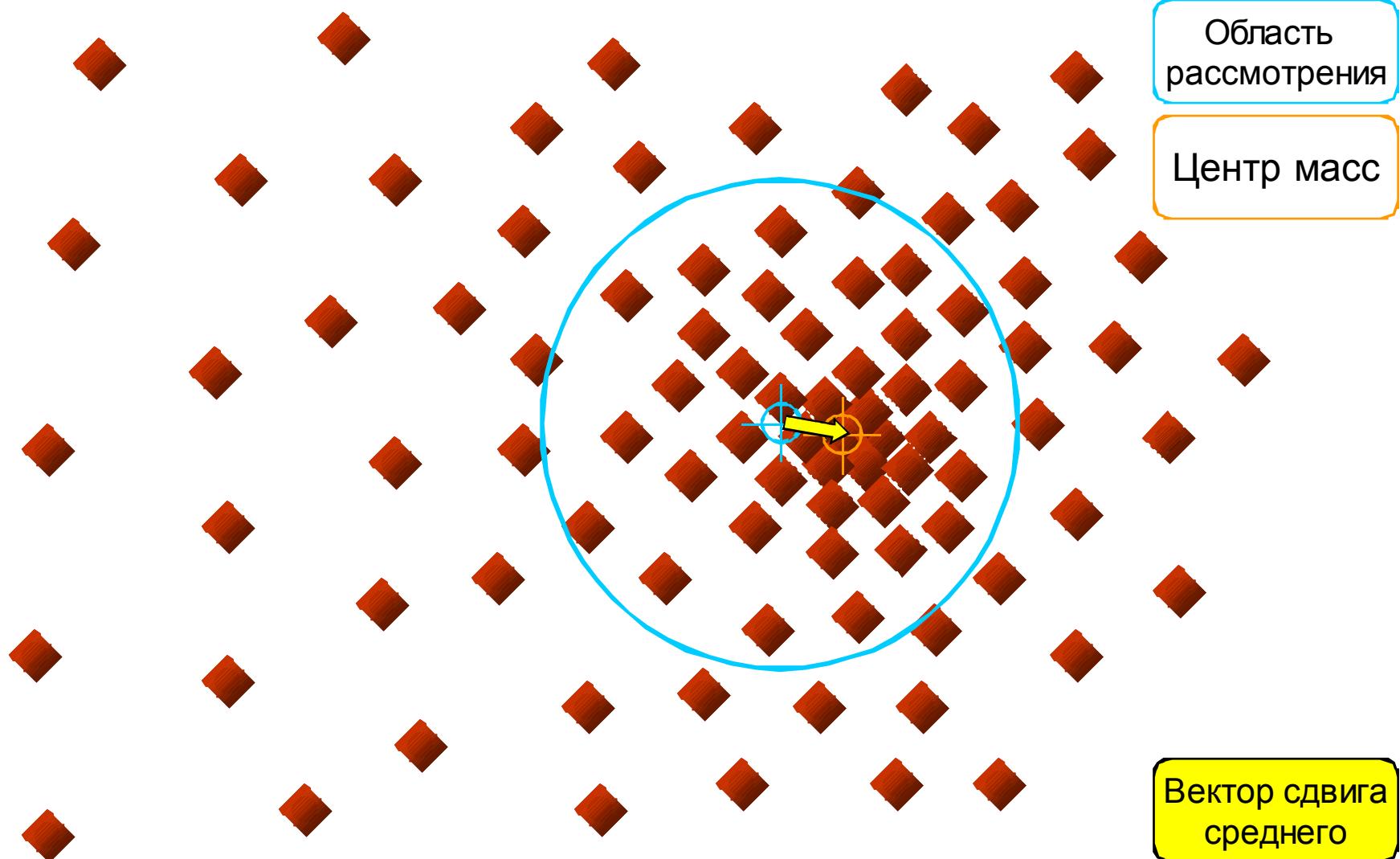
Метод сдвига среднего: общая идея



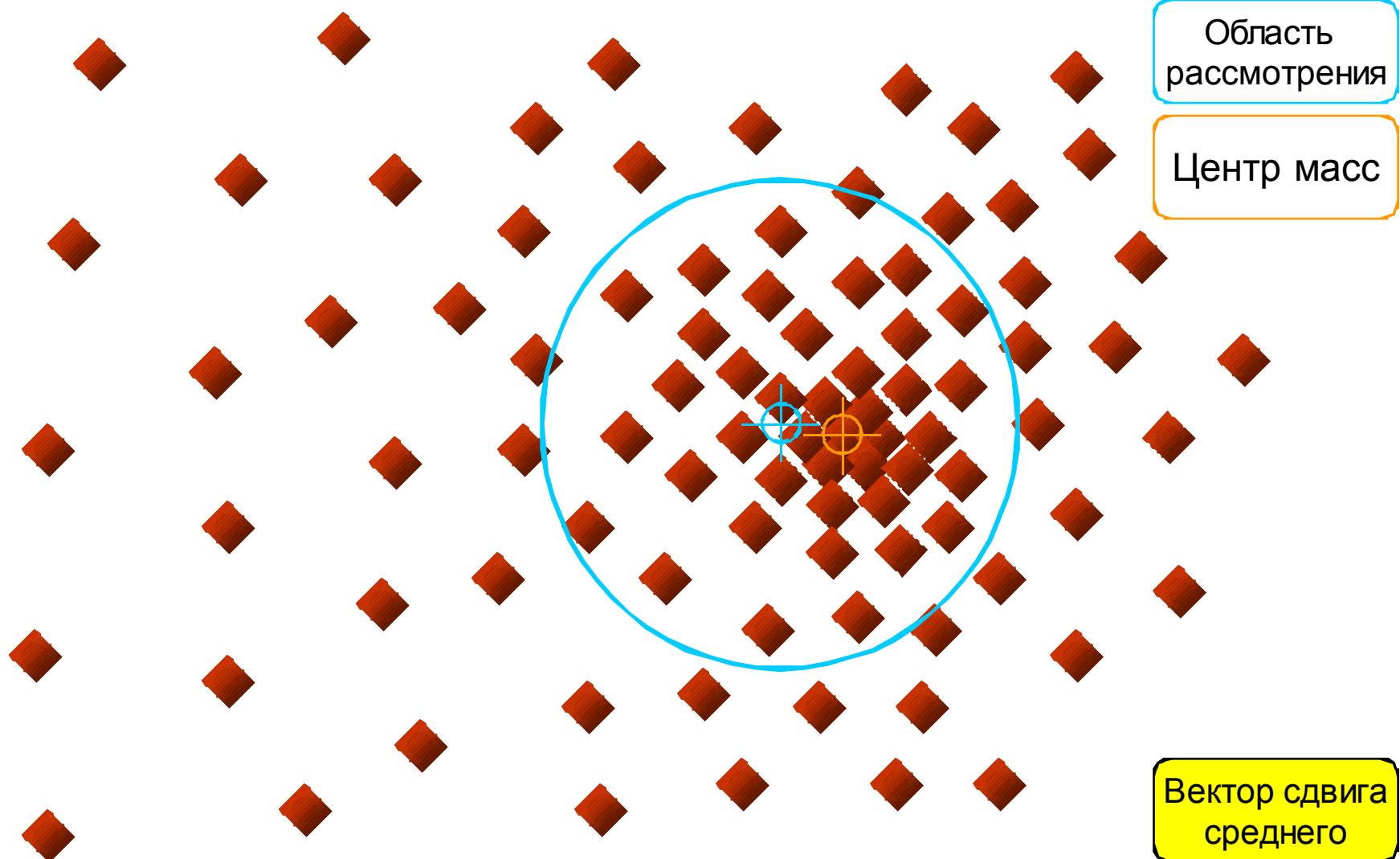
Метод сдвига среднего: общая идея



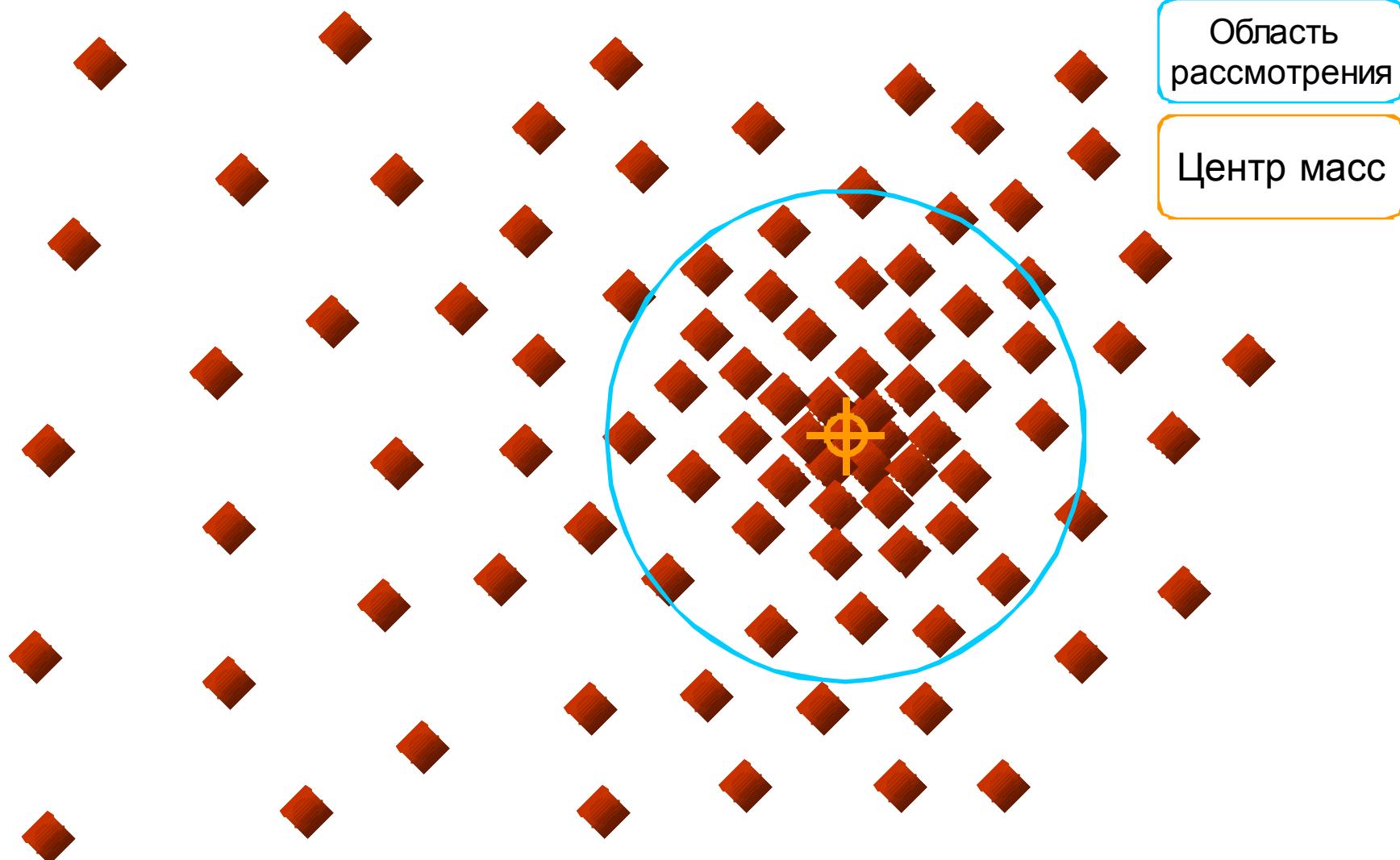
Метод сдвига среднего: общая идея



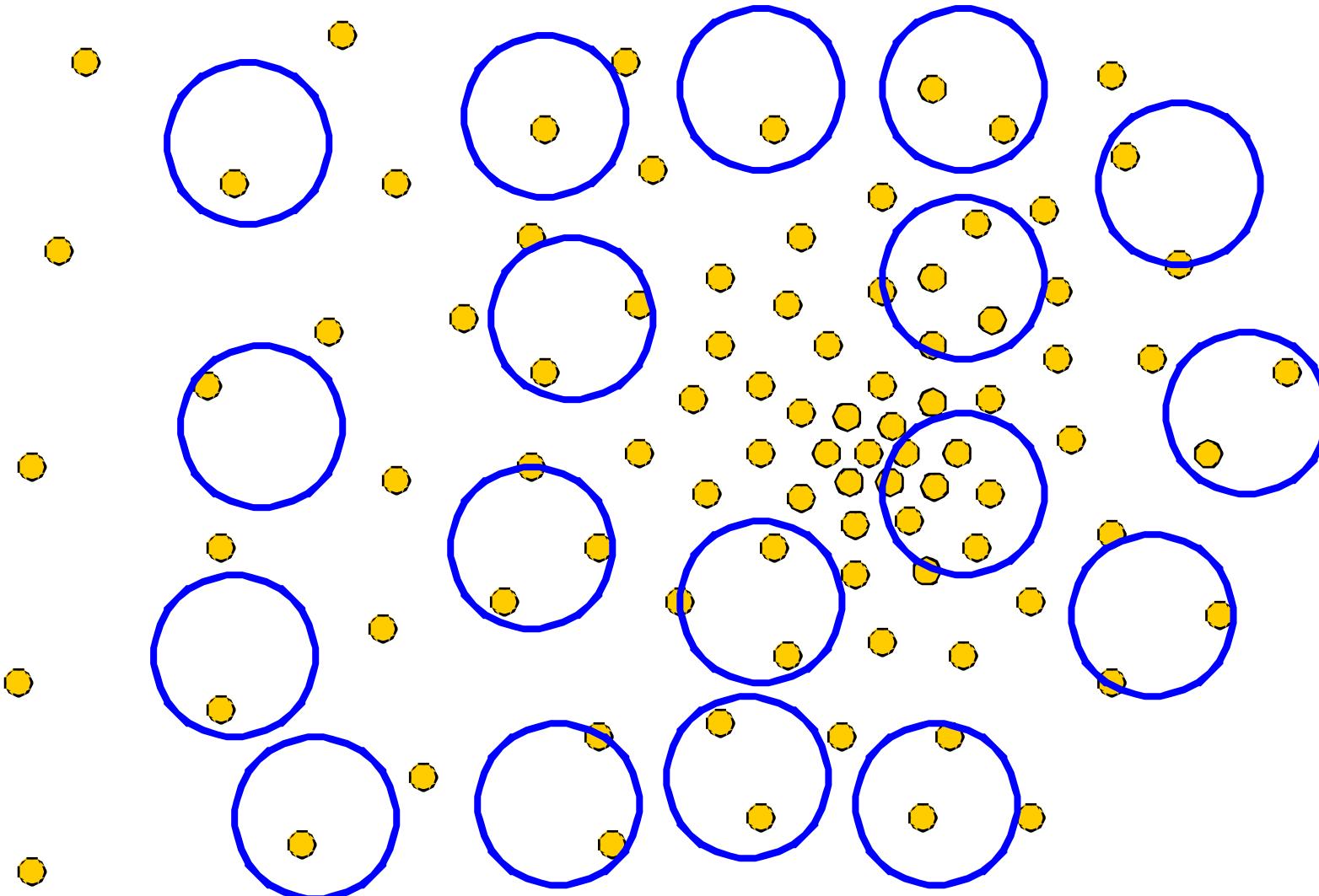
Метод сдвига среднего: общая идея



Метод сдвига среднего: общая идея



Вычисление мод распределения



Кластер: точки, сдвиг среднего для которых приводит к одной и той же моде распределения

Вычисление мод распределения

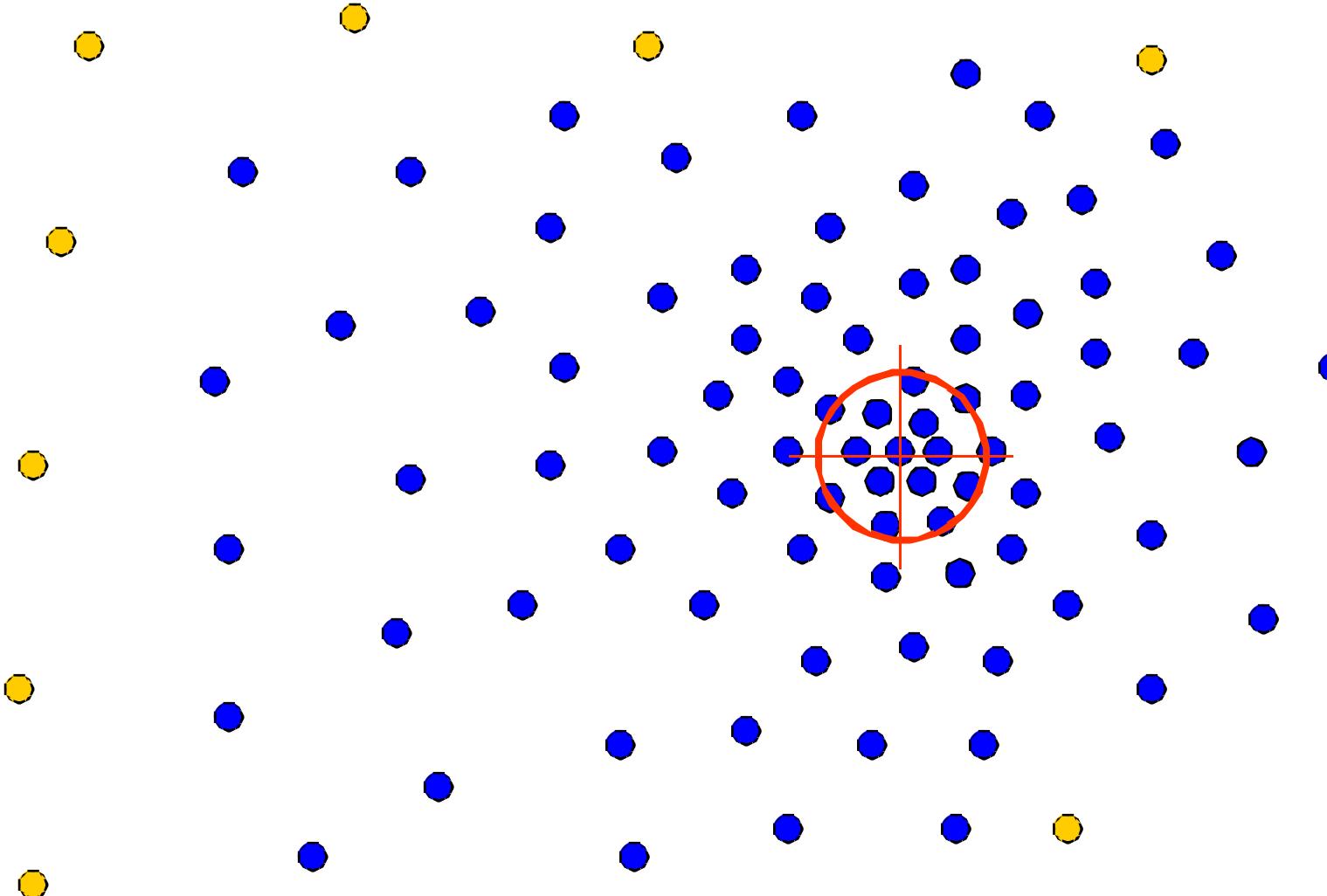
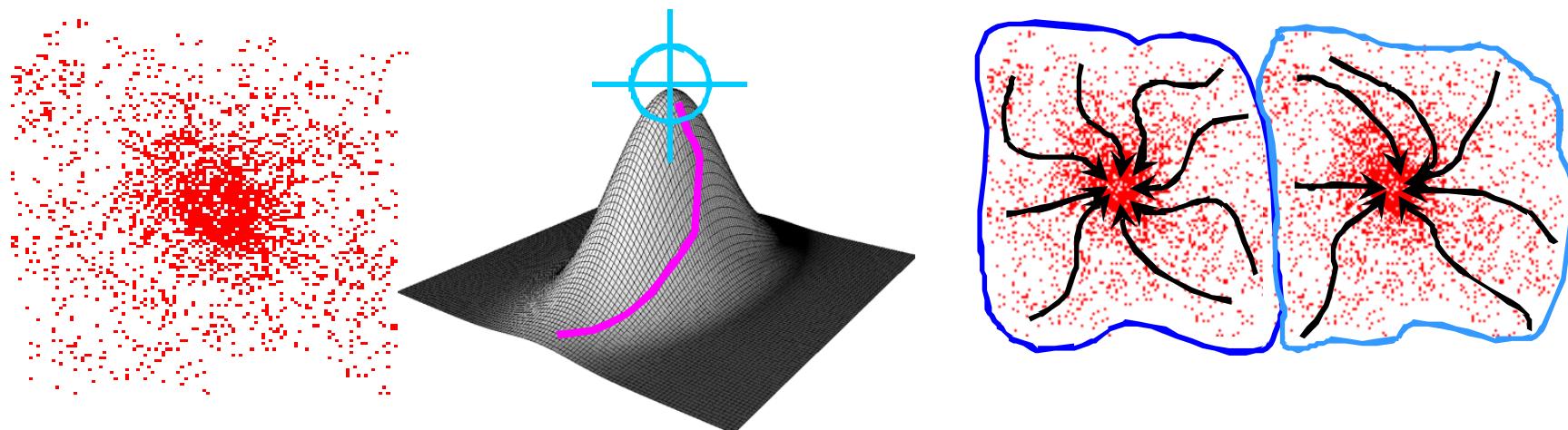


Схема метода



- Для каждой точки находим ближайшую моду плотности распределения (двигаясь вдоль градиента плотности)
- Разбиваем всю выборку на кластеры по ближайшим для них модам
 - Кластер – группа точек, для которых процедура поиска («сдвиг среднего») приводит к одной и той же моде распределения



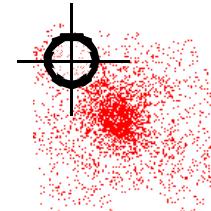
Ядерные методы (Kernel methods)



Для оценки плотности в MeanShift используются непараметрические методы – окна Парзена, основанные на функции ядра

$$P(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n K(\mathbf{x} - \mathbf{x}_i)$$

Функция конечного числа точек из выборки
 $x_1 \dots x_n$



данные

Свойства ядерных функций:

- нормализация

$$\int_{\mathbb{R}^d} K(\mathbf{x}) d\mathbf{x} = 1$$

- Симметрия

$$\int_{\mathbb{R}^d} \mathbf{x} K(\mathbf{x}) d\mathbf{x} = 0$$

- Экспоненциальное
убывание веса с расстоянием
до точки

$$\lim_{\|\mathbf{x}\| \rightarrow \infty} \|\mathbf{x}\|^d K(\mathbf{x}) = 0$$



Сдвиг среднего

- Возьмём градиент от оценки плотности распределения:

$$\nabla P(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \nabla K(\mathbf{x} - \mathbf{x}_i)$$

Не будем оценивать плотность-
оценим **градиент** плотности

- Можно показать, что вектор «сдвига среднего» совпадает с направлением градиента плотности

$$\nabla P(x) = \frac{c}{n} \sum_{i=1}^n \nabla k_i = \frac{c}{n} \left[\sum_{i=1}^n \nabla g_i \right] \left[\frac{\sum_{i=1}^n x_i g_i}{\sum_{i=1}^n g_i} - x \right] \quad g(x) = -k'(x)$$

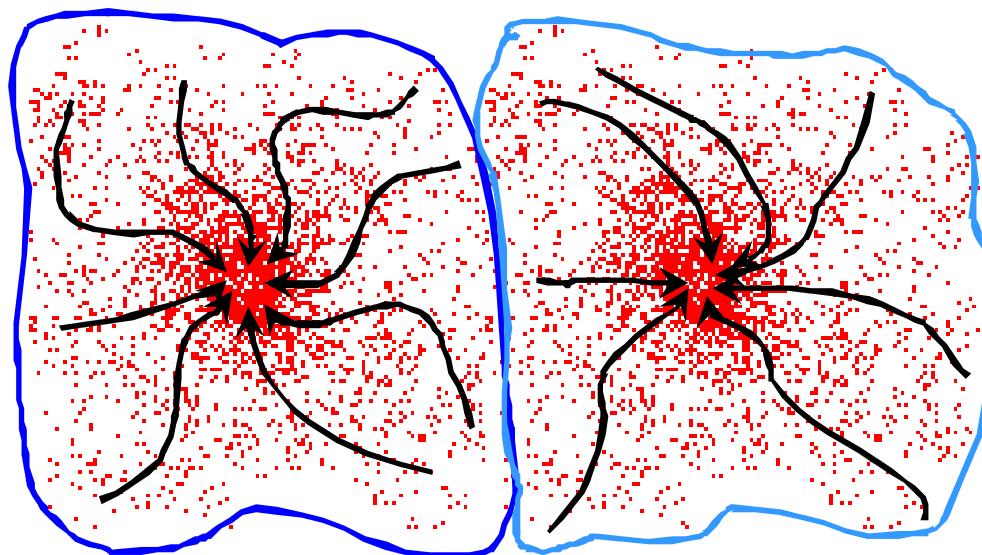
$$\mathbf{m}(\mathbf{x}) = \left[\frac{\sum_{i=1}^n \mathbf{x}_i g\left(\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{h}\right)}{\sum_{i=1}^n g\left(\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{h}\right)} - \mathbf{x} \right]$$

«сдвиг среднего» - вектор между текущим положением и взвешенным среднем

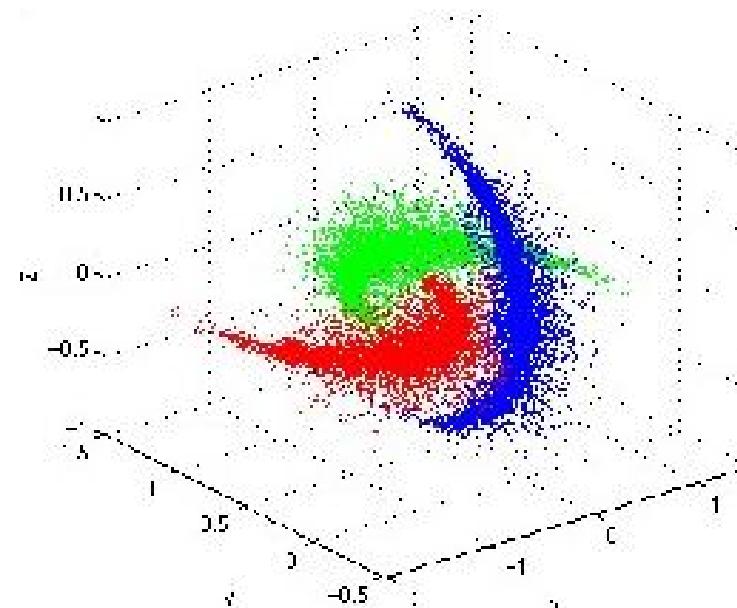
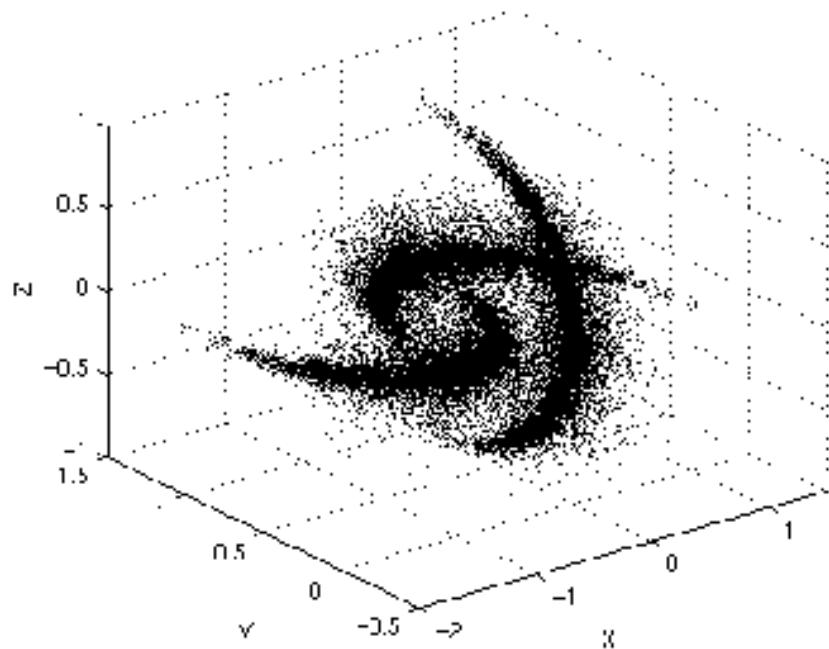
Кластеризация



- Сложность метода $O(dN^2T)$, где N – число точек, T – число итераций алгоритма. $dT \ll N$
 - Рассчитывает траекторию для каждой точки N
 - Оценка плотности пропорциональна N



Пример работы и свойства

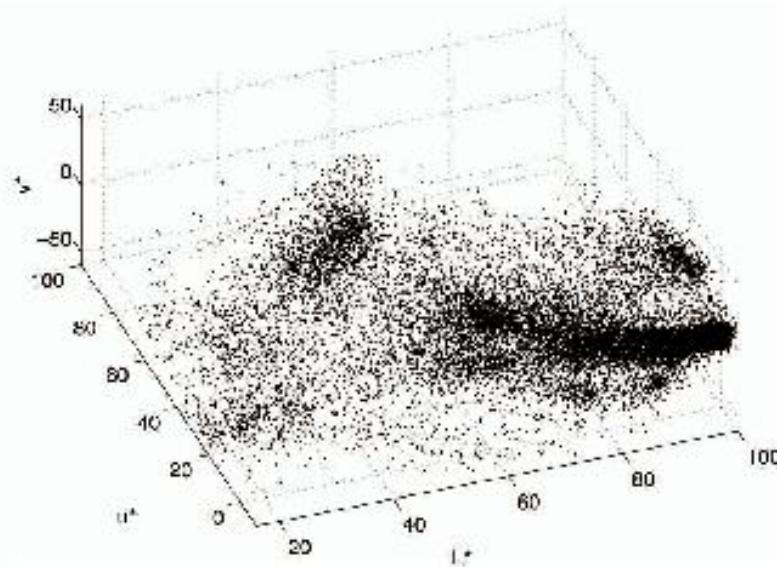


- (+) Не делается предположений о форме кластеров и форме распределения
 - MeanShift позволяет учитывать сложную форму кластеров
- (+) Число кластеров определяется автоматически
- (-) Выбор параметров ядровой функции нетривиален
 - От них сильно зависит работа

Применение для изображений



Кластеризация по цвету:

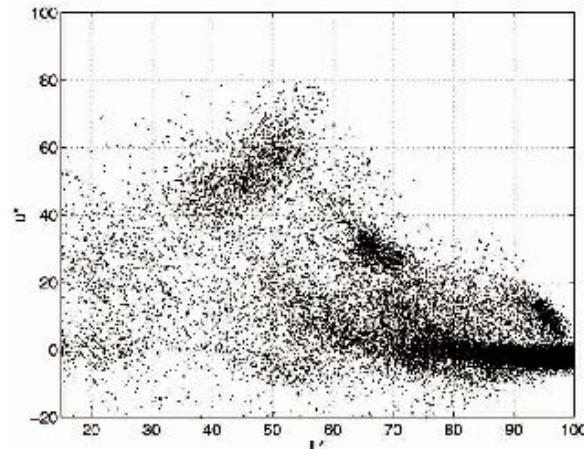


Точки в пространстве $L^*u^*v^*$

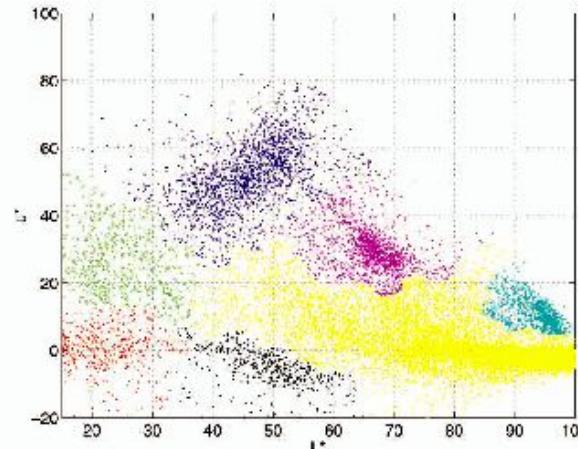
Применение для изображений



Точки в
пространстве
(L^*u)

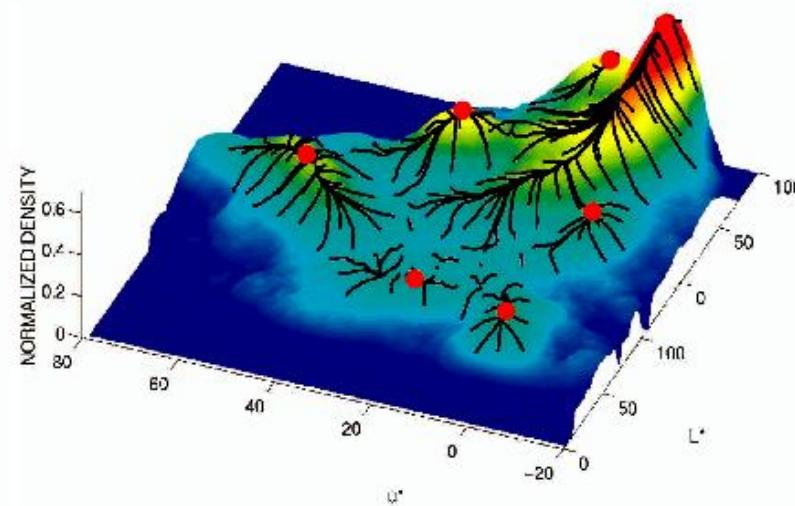


(a)



(b)

Полученные
кластеры

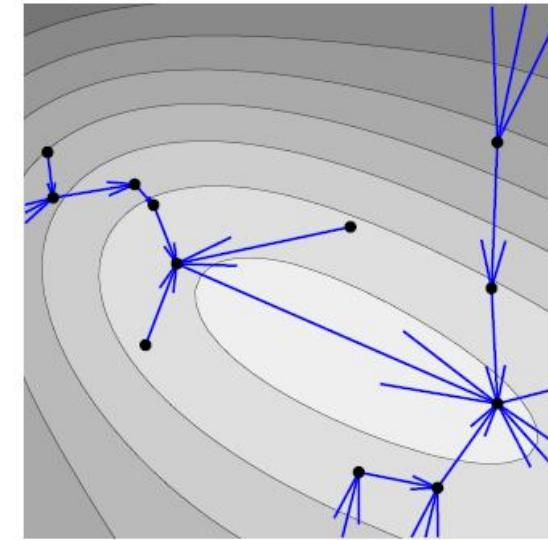
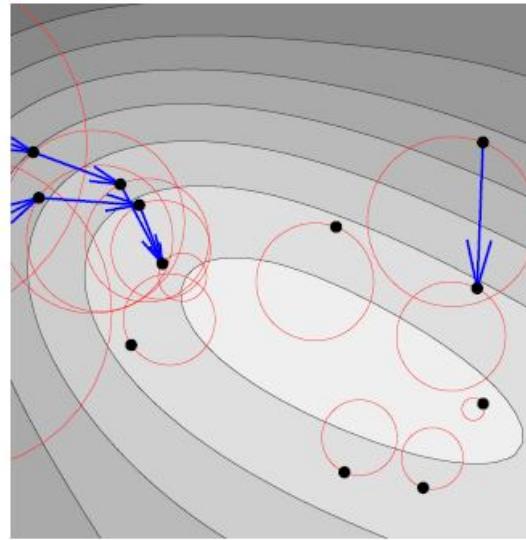
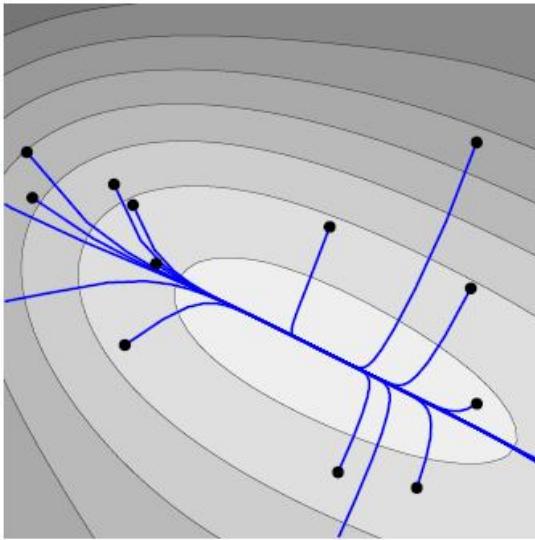


(c)

Пример сегментации



QuickShift



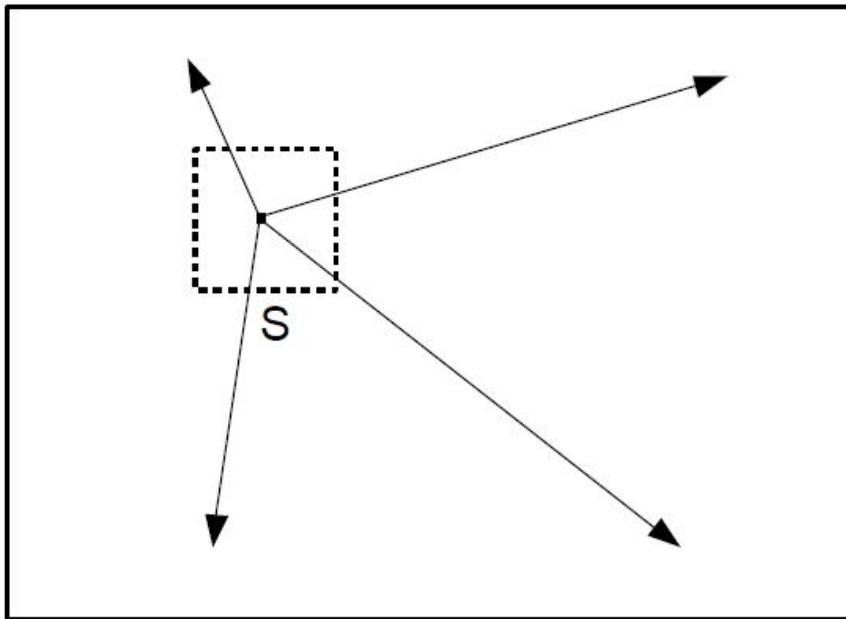
- MedoidShift Пусть наши траектории будут проходить только через точки выборки
- QuickShift – найдём для каждой точки ближайшего соседа, у которого плотность P больше

$$y_i(1) = \underset{j: P_j > P_i}{\operatorname{argmin}} D_{ij}, \quad P_i = \frac{1}{N} \sum_{j=1}^N \phi(D_{ij}), \quad D_{kj} = d^2(x_k, x_j)$$

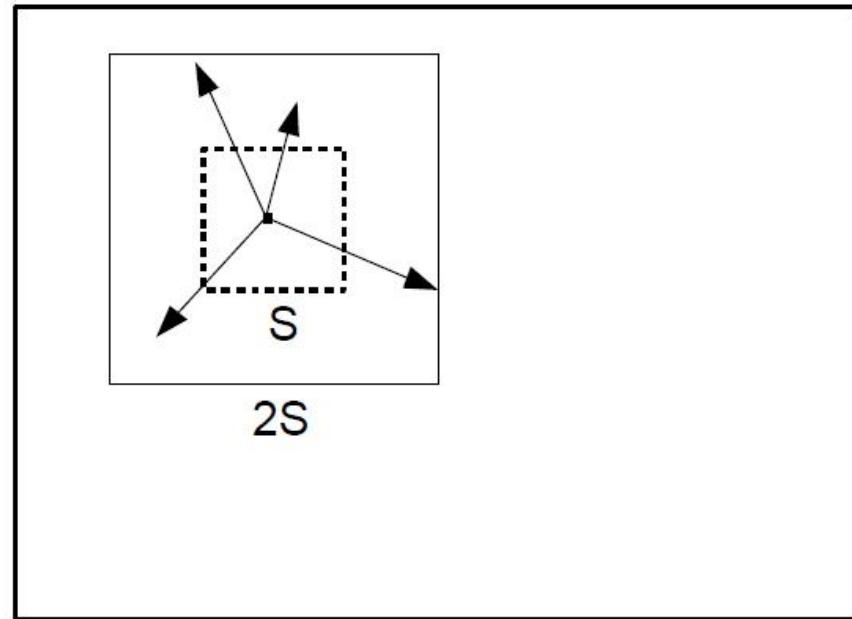
- Фактически, объединим все точки в дерево. Затем будем «рвать» самые длинные рёбра
- Сложность $O(dN^2)$ где d – размерность данных

A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In Proc. ECCV, 2008.

Simple linear iterative clustering (SLIC)



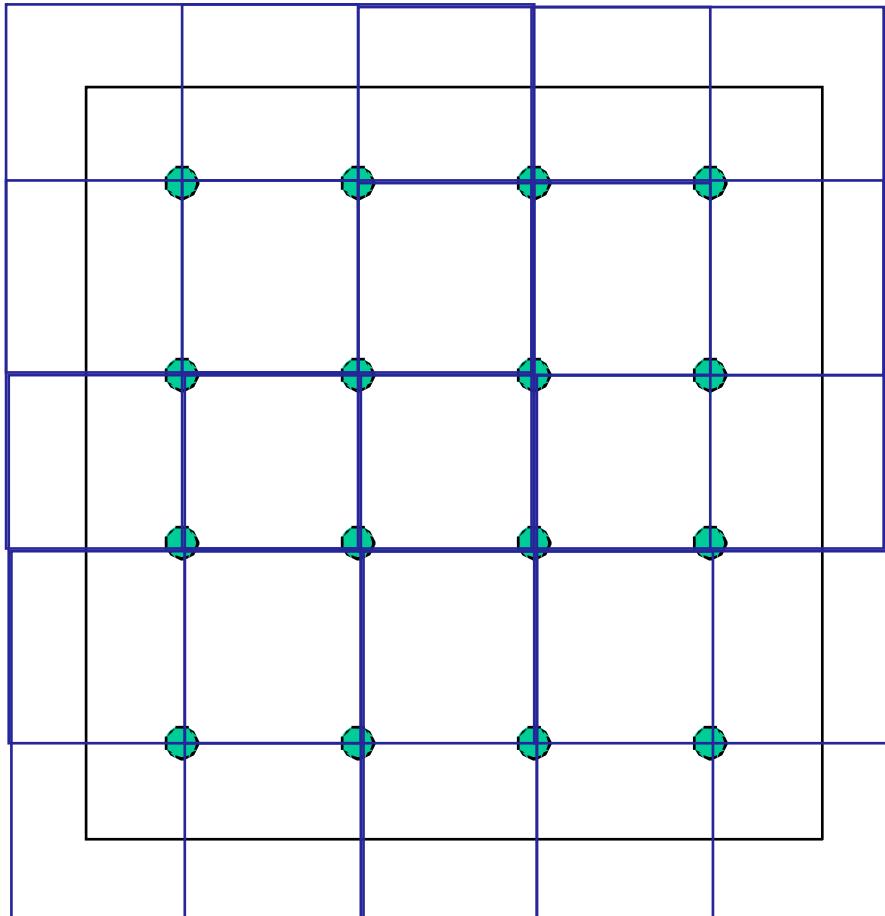
(a) standard k -means searches
the entire image



(b) SLIC searches
a limited region

- Зачем сопоставлять пиксель со *всеми* остальными?
- Инициализируем кластеры по сетке на расстоянии s
- Будем сравнивать с «центром кластера» только пиксели, которые *могут* принадлежать этому сегменту (на расстоянии $< s$)

Алгоритм SLIC



- Инициализируем центры кластеров C_k по сетке с шагом S
- Инициализируем в каждом пикселе метку $L(i) = -1$ и расстояние $D(i)$ до ближайшего кластера $-\infty$
- Проходим по кластерам C_k
 - Для каждого пикселя в области $2S \times 2S$ считаем расстояние до C_k
 - Если расстояние меньше $D(i)$, тогда ставим метку $L(i) = k$, и записываем в $D(i)$ новое значение
- Повторяем до тех пор, пока суммарное изменение кластеров не будет меньше порога

Функция сравнения



- Работаем в LABCIE + (x,y) пространстве
- Суммируем расстояния по цвету и положению и нормализуем на максимальное расстояние внутри кластера

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2}$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

$$D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2}.$$

- Поскольку расстояние по цвету неизвестно, можем зафиксировать его как m

$$D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \quad \rightarrow \quad D = \sqrt{d_c^2 + \left(\frac{d_s}{S}\right)^2 m^2}$$

Примеры работы



Сложность - $\left(\frac{n}{S}\right)^2 S^2 t = n^2 t = Nt$

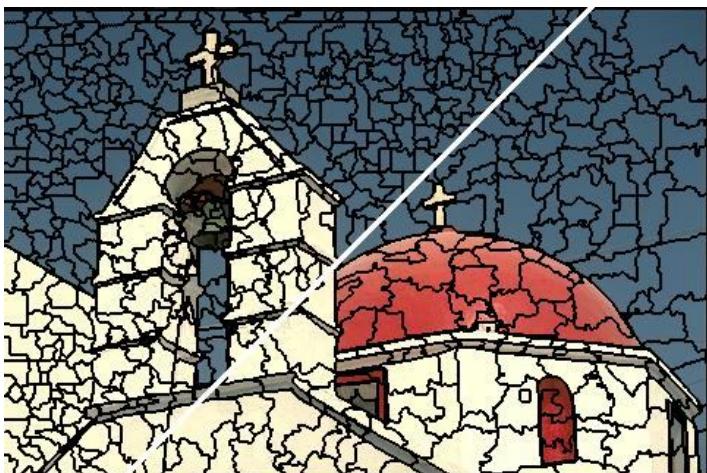
Визуальное сравнение



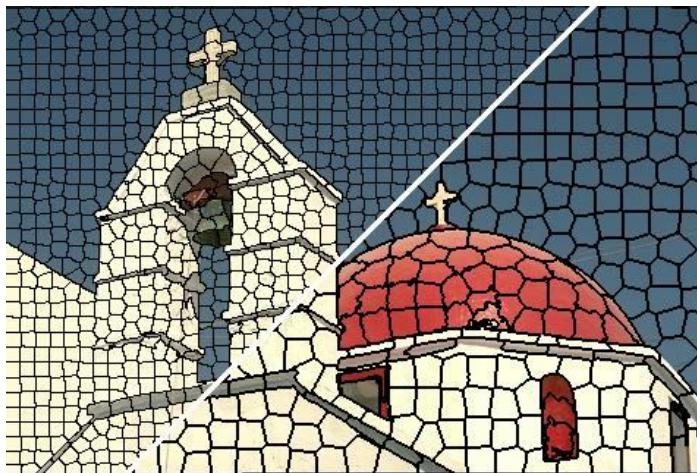
Efficient Graph-Based



TurboPixel

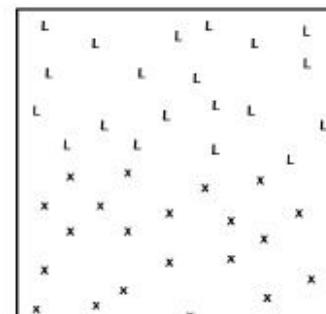
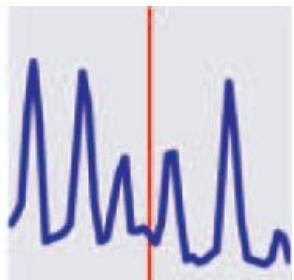
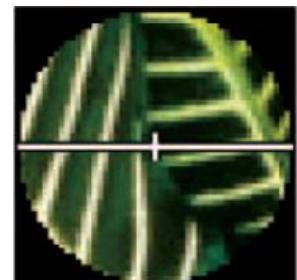
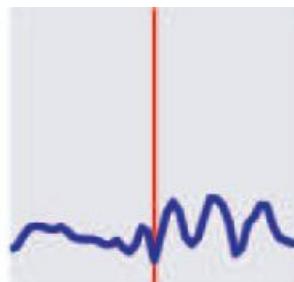
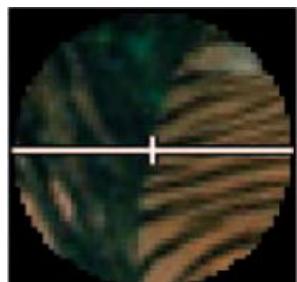
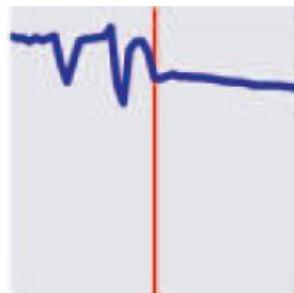
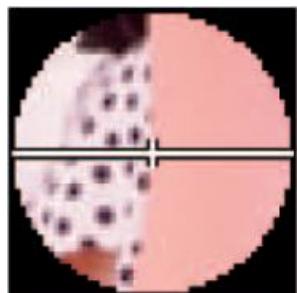


QuickShift

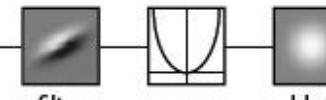


SLIC

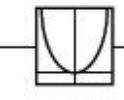
Учёт текстуры?



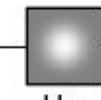
input



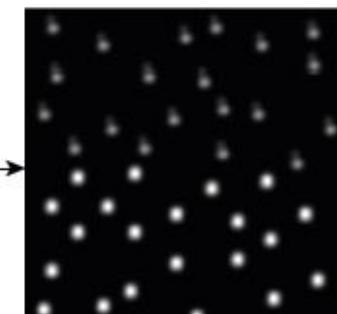
filter



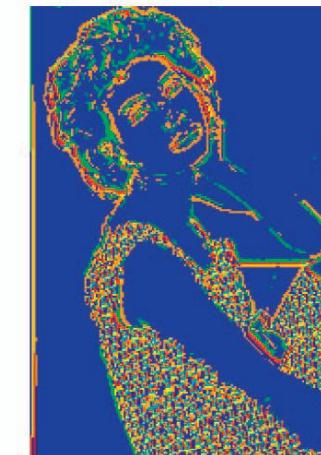
square



blur



output



- В любых методах пересегментации можно добавить учёт текстуры
- Признаки на основе отклика банка фильтров
 - Напрямую отклики фильтров
 - Квантование по словарю («текстоны», аналог мешка слов, но для текстур)

Промежуточное резюме



- Сегментация изображения на «суперпиксели» (локализованные области, однородные по своим характеристикам) может использоваться как предобработка для решения других задач
 - Понижение размерности задач (работа с областями, а не с отдельными пикселями)
 - Хорошая сегментация должна учитывать несколько признаков в совокупности – цвет, текстуру
- Мы рассмотрели ряд методов на основе кластеризации:
 - Сдвиг среднего и производные
 - SLIC
 - Описание текстуры с помощью текстонов и применения для поиска краёв



Семантическая сегментация

Использование CRF

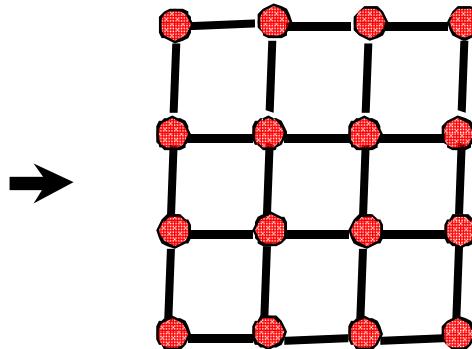


Простой вариант учёта локального контекста – использовать CRF модель из интерактивной сегментации

Изображение



CRF



Парные члены



$$Energy_{MRF} = E(F) = \sum_p (\phi(I | F_p)) + \sum_{\{p,q\} \in N} (\phi(I | F_p, F_q) * \psi(F_p, F_q))) + const$$

Унарное правдоподобие (Unary likelihood) Член контраста (Contrast Term) Модель априорной вероятности Поттса (Potts Model Prior)

$$MAP \ solution \quad F^* = \arg \min_F E(F)$$

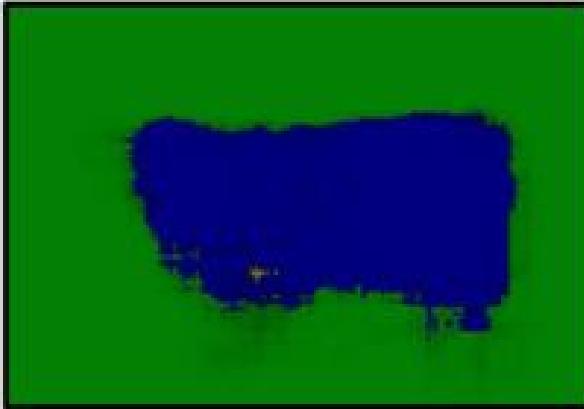
Вклад CRF



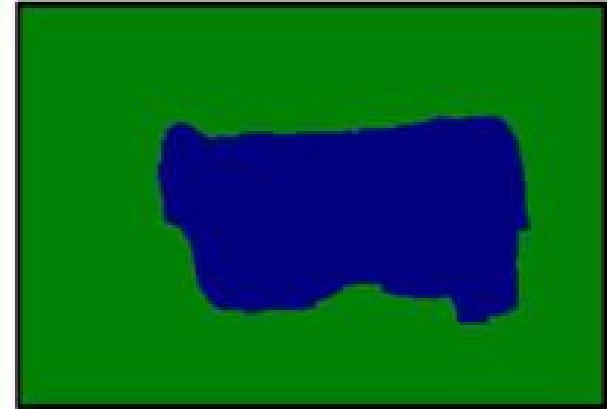
Пример из работы TextonBoost



Исходное изображение



Независимая
попиксельная
классификация

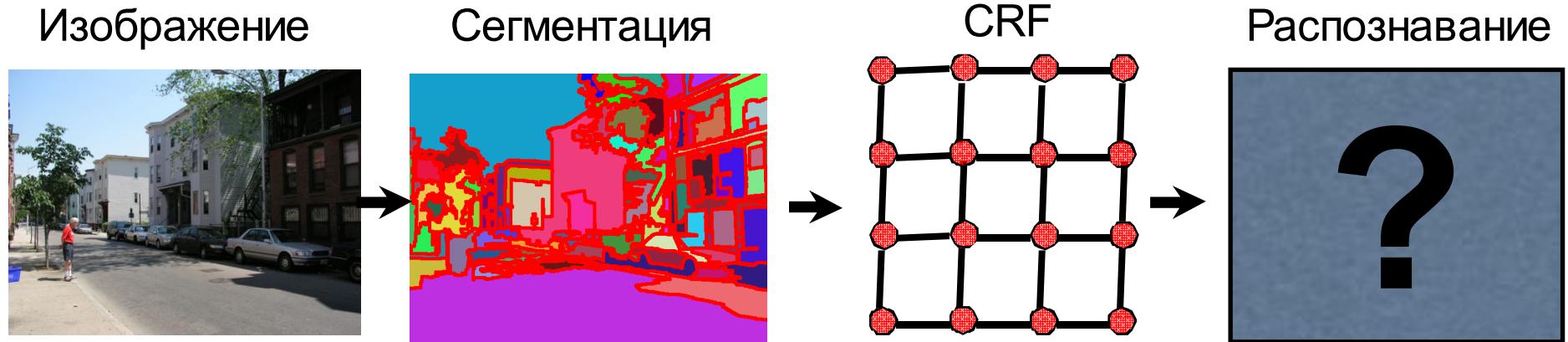


Сглаживание CRF

- При использовании только визуального сходства в парных потенциалах
- Вклад CRF – сглаживание
- Наблюдается «shrinking bias»

J. Shotton, J. Winn, C. Rother, A. Criminisi, *TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation*, ECCV 2006

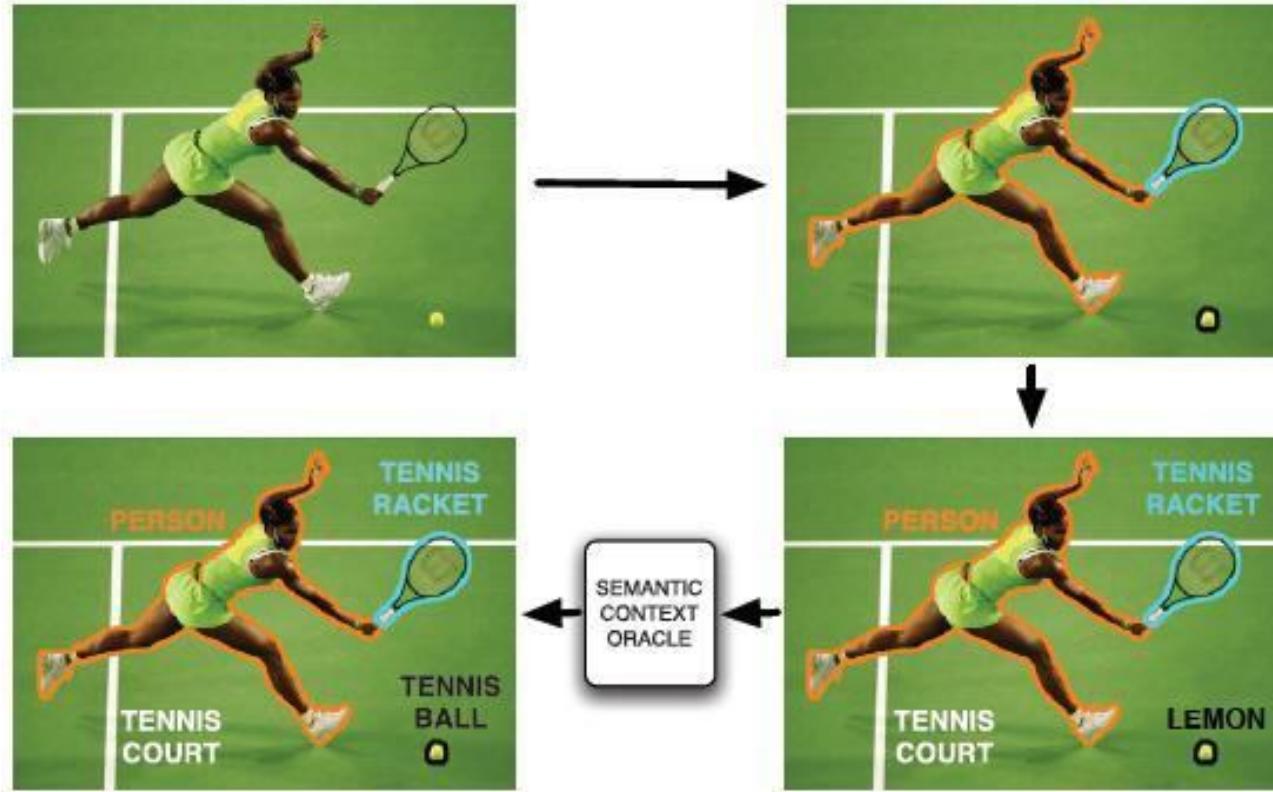
CRF на суперпикселях



- Сегментация изображения на суперпиксели (QuickShift)
- Задание CRF (условного случайного поля) на графе смежности на суперпикселях
- Разметка - вывод в CRF с помощью альфа-расширения
 - Оптимизация «энергии»

Brian Fulkerson, Andrea Vedaldi, Stefano Soatto. Class Segmentation and Object Localization with Superpixel Neighborhoods, CVPR 2009

Объекты в контексте



- Некоторые пары соседства объектов встречаются реже, чем другие
- Пример расширения функций «парных потенциалов»

Andrew Rabinovich, Andrea Vedaldi, Carolina Galleguillos, Eric Wiewiora,
Serge Belongie, Objects in Context, ICCV2007

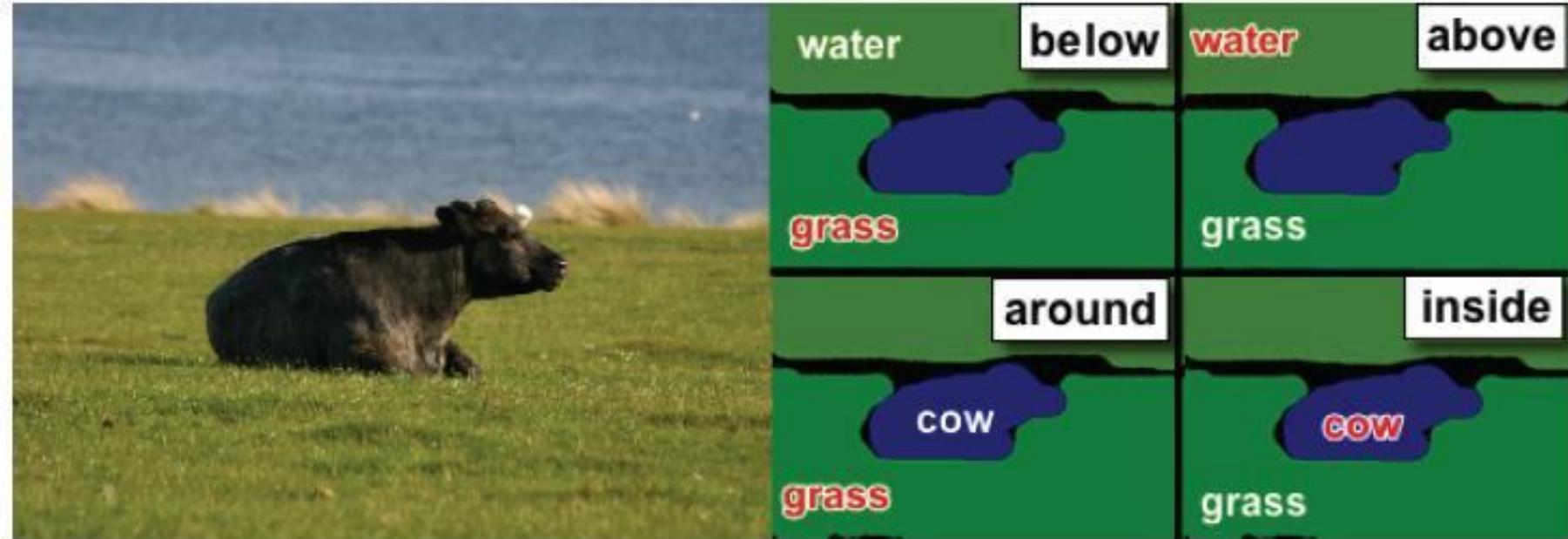
«Со-occurrence matrix»



MSRC training data																				
building	grass	tree	cow	sheep	sky	aeroplane	water	face	car	bike	flower	sign	bird	book	chair	road	cat	dog	body	boat
75 18 29	33 6 9 7 18 10	2 1	43	1 9 8																
18 93 38 23 15 39	14 7 7	3 1	1	4 15	2 8															
29 38 68 6	43 6 12 9 4	1 2	1 19	11 8																
23 6 23	4 4																			
15	15	1																		
33 39 43 4	86 15 18 4 3	5 4	25	8 11																
6 14 6	15 15		5																	
9 7 12 4 1	18 43 4 1	7	6	6 18																
7 7 9	4 4 28 1 1 1	3	7	7 28 1																
18 4	3 1 1 20		19	1																
10 3	1 15		12	1																
1 1	1 1																			
2 2	5																			
1 1	4 7																			
43 15 19	2 25 5 6 7 19 12	1 3	3 86	7 10 8 1																
cat			7 7																	
dog			10 13																	
body			8																	
boat			1 1 2 19																	

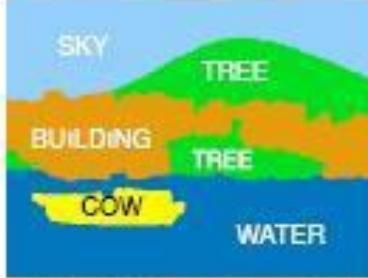
- Матрица совместного появления объектов
- Диагональные элементы – частоты появления объектов в обучающей выборке
- Внедиагональные элементы = количество совместных появлений объектов
- Специальный парный штраф $\Phi(c_i, c_j)$ можно обучить по этой таблицы

Взаимное расположение объектов



- Можно дополнительно штрафовать маловероятные комбинации взаимного расположения объектов
- Тоже можно обучить по тренировочной выборке
- На рисунке несколько примеров возможных расположений

Несколько примеров

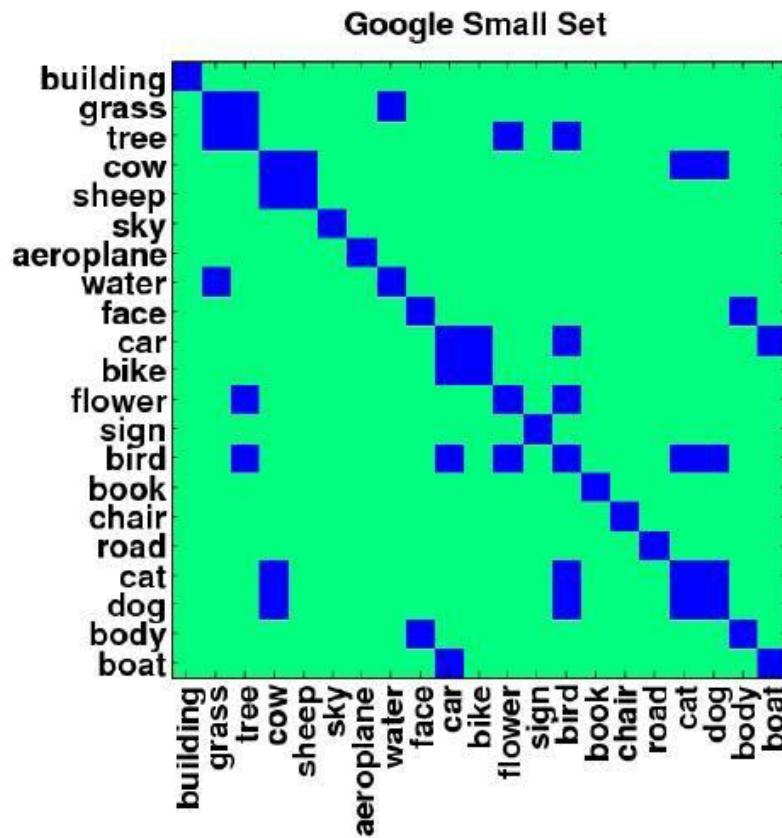


Некоторые глупые ошибки можно победить таким способом

Помощь «интернета»



- Некоторую информацию можно автоматически собирать в интернете
 - Например, воспользуемся поиском для оценки частоты совместного появления объектов



Сравнение



	No Context	Google Sets	Using Training
MSRC	45.0%	58.1%	68.4%
PASCAL	61.8%	63.4%	74.2%

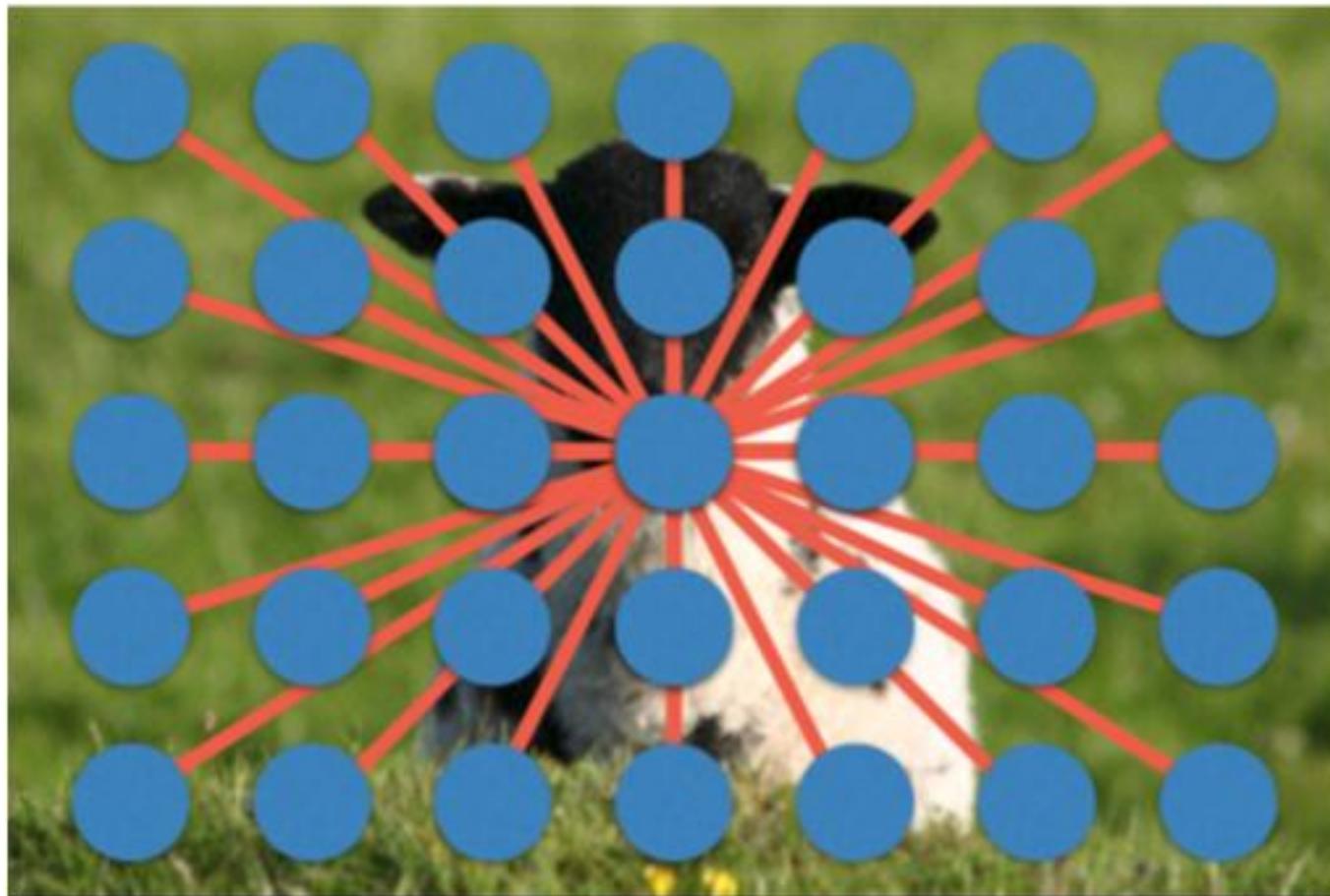
Table 1. Average Categorization Accuracy.

- Сравнение совместной встречаемости объектов, вычисленное через запросы в Google и разметку
- Видно, что качественная эталонная разметка всё-таки пока лучше очень шумной автоматической оценки

Плотносвязанные CRF



$$E(\mathbf{x}) = \sum_i \phi_i(\mathbf{x}_i) + \sum_{i,j \in \mathbb{N}} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$



Плотносвязанные CRF



- Bilateral Pairwise Term

$$\psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \mu(\mathbf{x}_i, \mathbf{x}_j) \exp\left(-\frac{|\mathbf{s}_i - \mathbf{s}_j|^2}{2\sigma_\alpha^2} - \frac{|\mathbf{c}_i - \mathbf{c}_j|^2}{2\sigma_\beta^2}\right)$$

Label compatibility

Location smoothness

Appearance smoothness

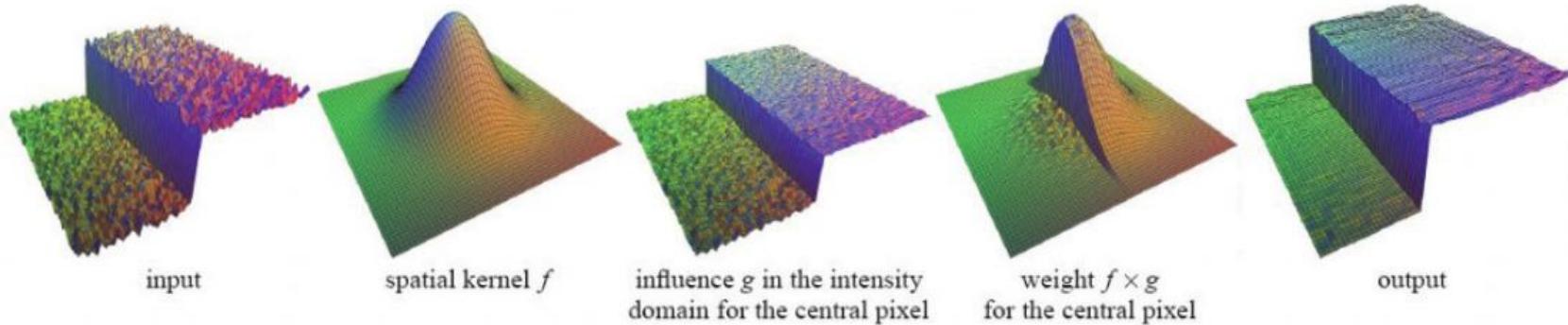


Image credit: Krähenbühl et al.

Плотносвязанные CRF



- Дальние связи позволяют преодолеть «shinking bias»
- Огромное количество связей, как делать вывод?
 - Подробнее рассмотрено на семинаре!
- Активно используется сейчас

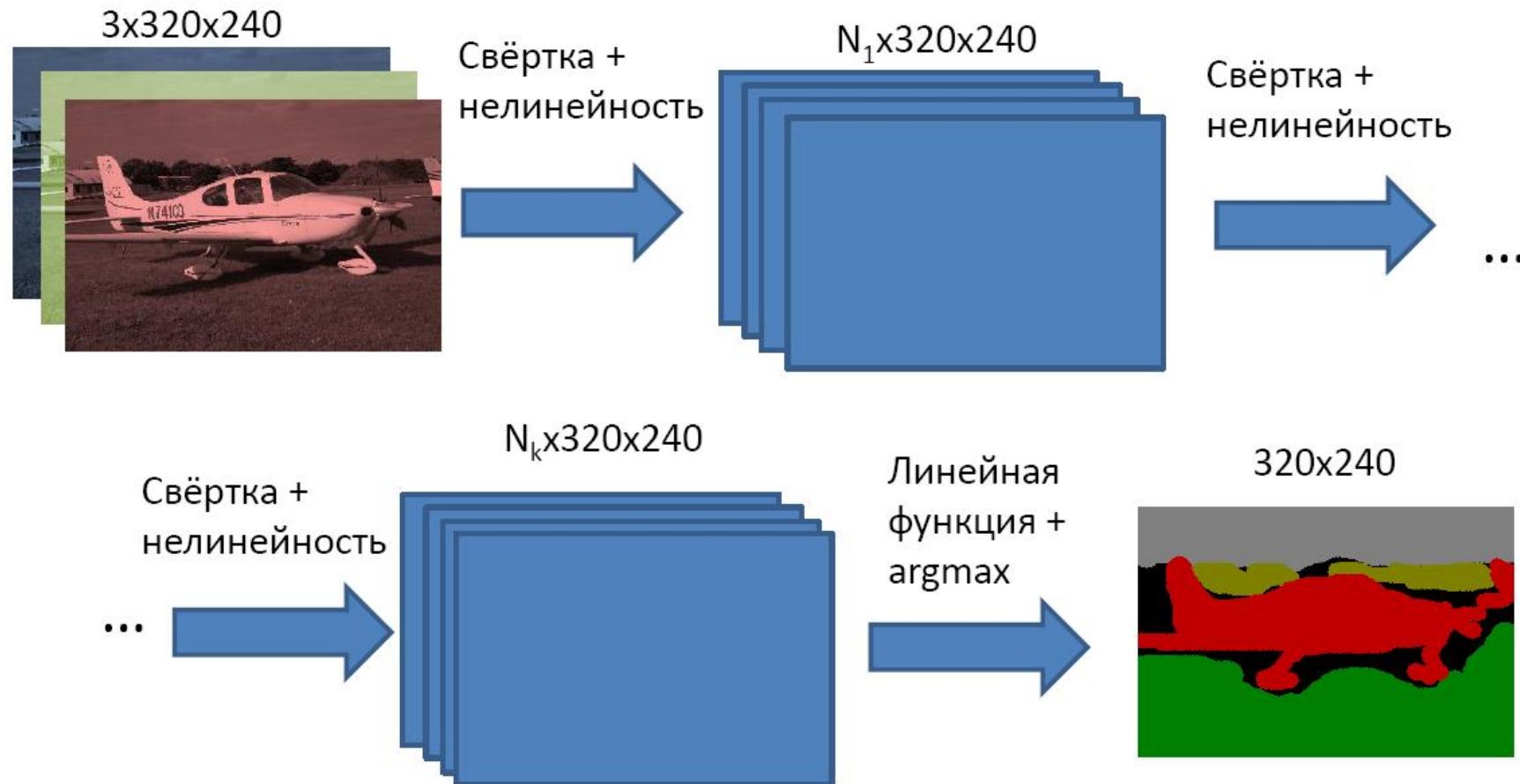


Image credit: Krähenbühl et al.



Нейросетевые модели

Нейросетевая модель

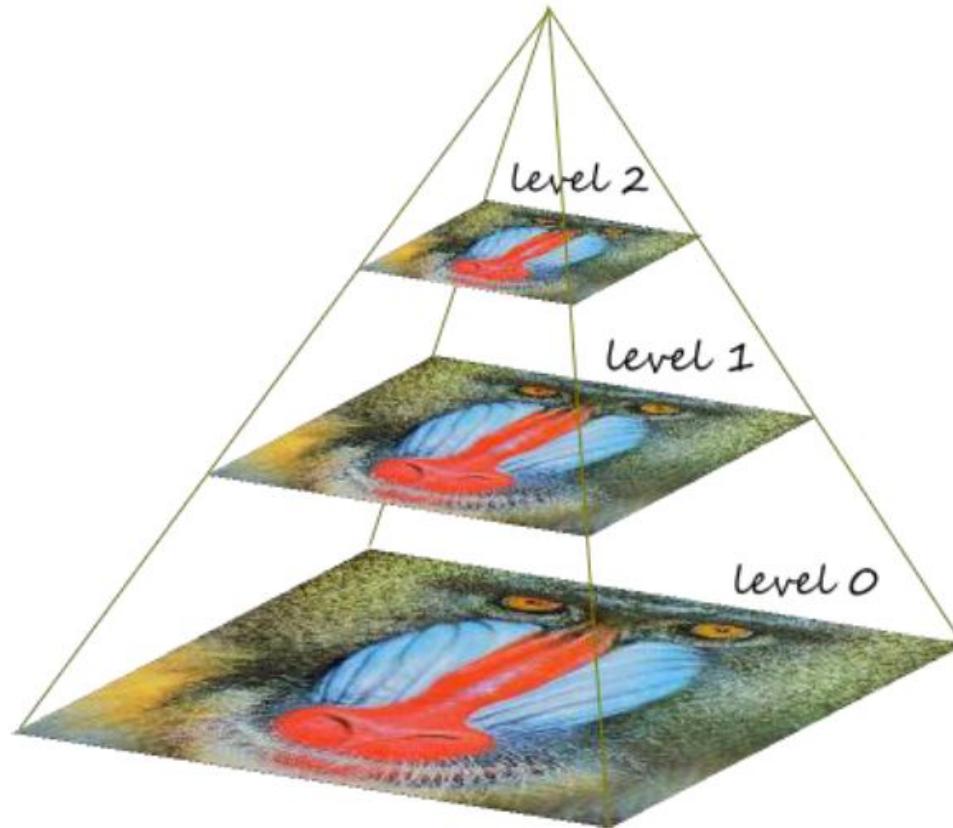


- Упаковываем контекст внутрь нейросетевой модели
- Это простейший пример

Инвариантность к масштабу

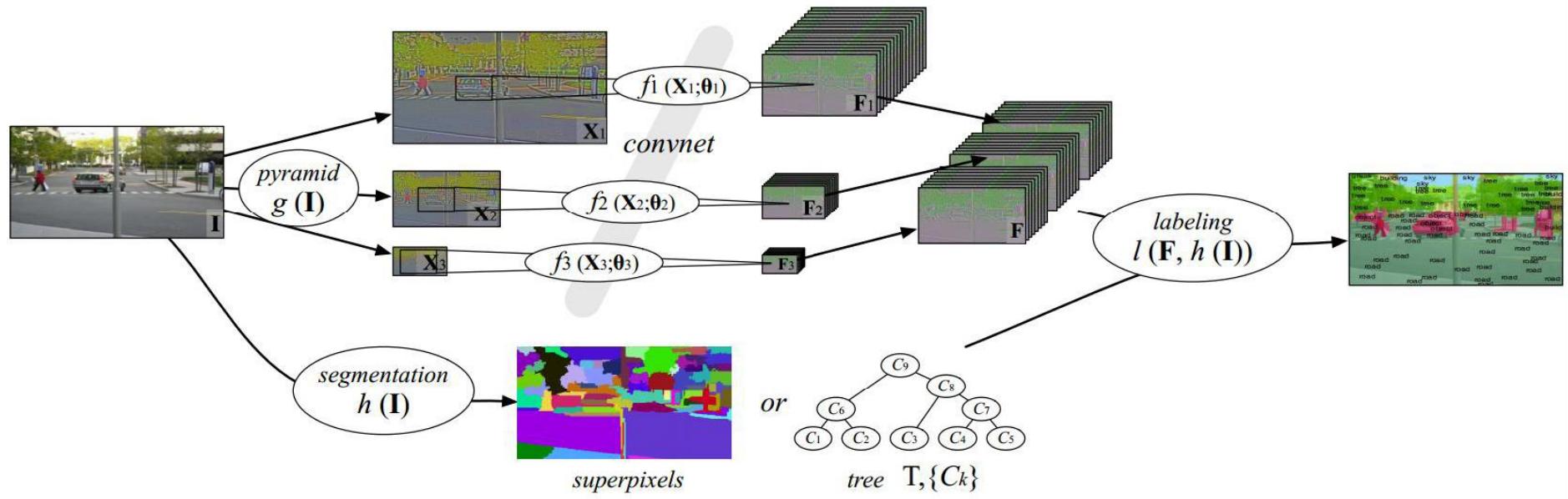


Пирамида изображений



Подсчитаем **одинаковые** признаки по разным уровням пирамиды.

Многомасштабные признаки



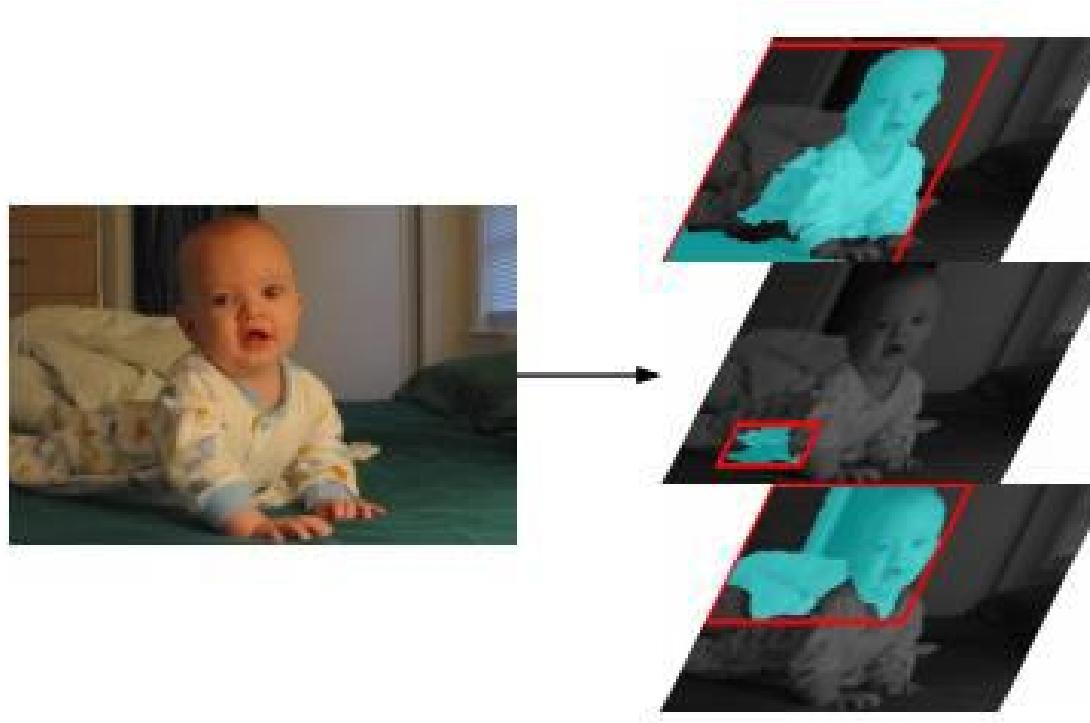
- Попиксельный нейросетевой классификатор
- Объединение выходов в рамках суперпикселей
- Др. вариант – построение CRF

Farabet, Learning Hierarchical Features for Scene Labeling, 2013

Классификация регионов



- Попробуем одновременно решать задачу и выделения и сегментации объектов
 - Применяем метод сегментации для генерации гипотез
 - Классификация гипотез
 - Уточнение сегментации

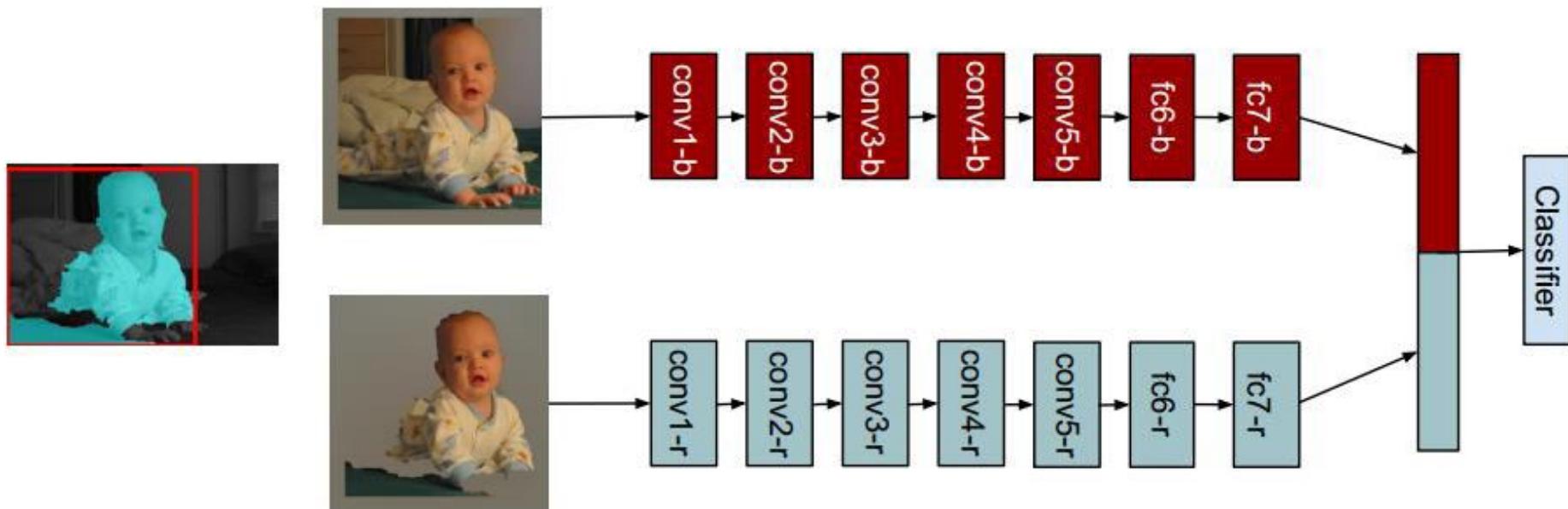


Hariharan, Simultaneous Detection and Segmentation, 2014

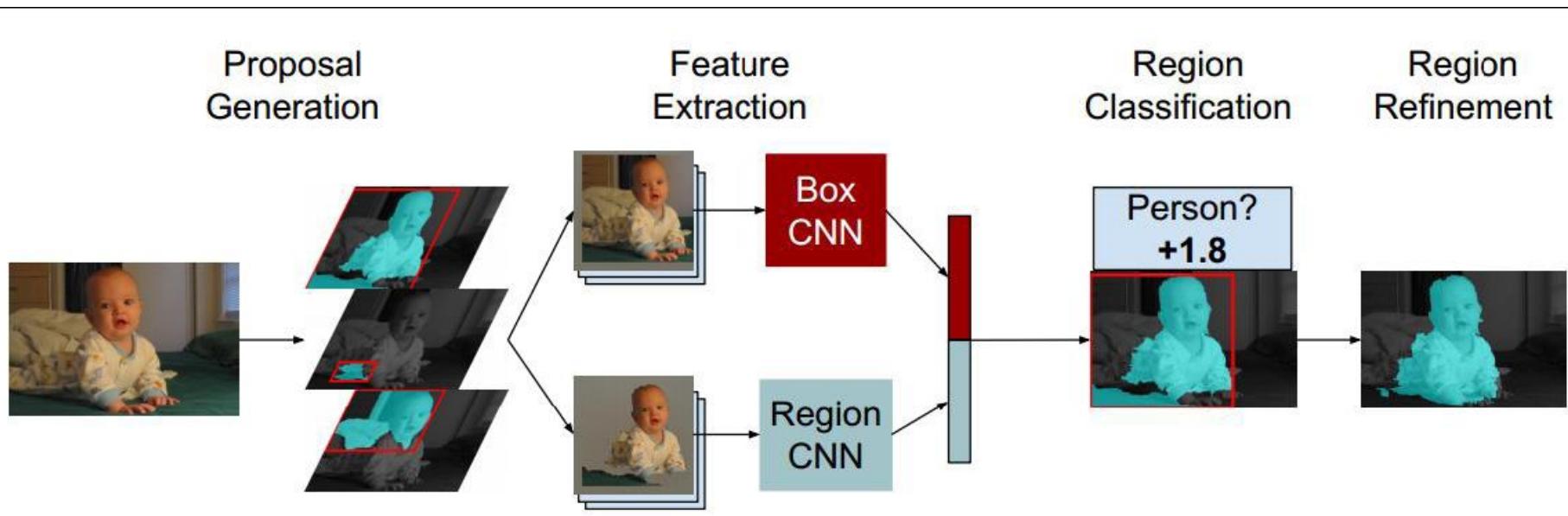
Классификация регионов – обучение



- Используем сеть, обученную на ImageNet
- Дообучим её двумя способами:
 - на предложенных патчах
 - на выделенных регионах внутри патча



Классификация регионов - тестирование



Результаты

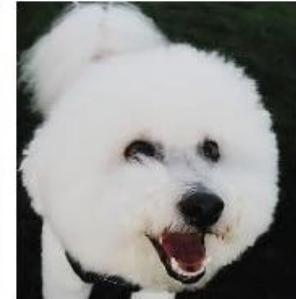


- Находит отдельные объекты
- Высокое качество

Слабая разметка



Кошки



37345 изображений

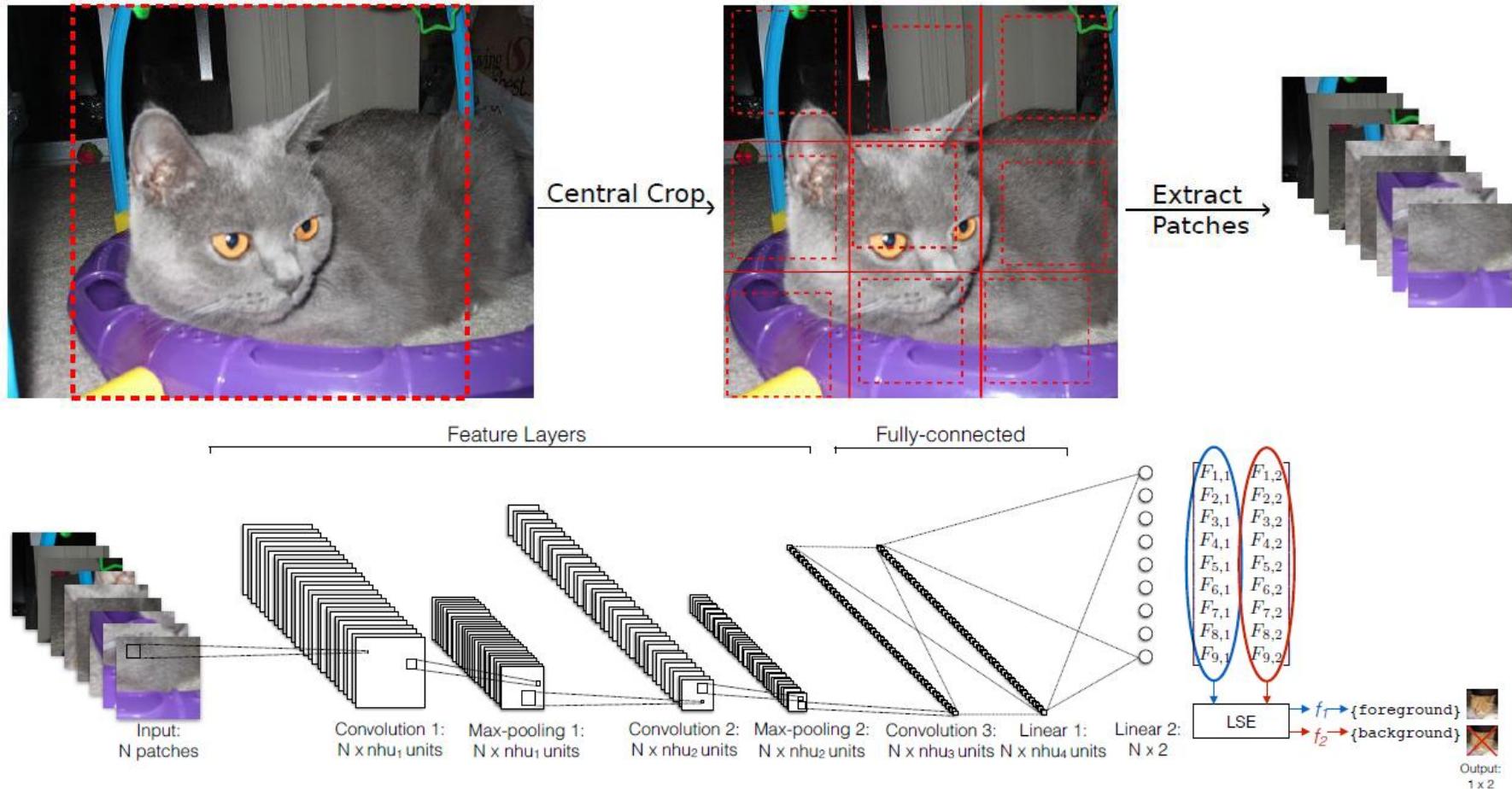
187775 изображений

- Можем ли мы зная только о наличии / отсутствии объекта на изображении научиться его сегментировать?
- «Слабая разметка» - меньшая информация по сравнению с целевой

IMAGENET

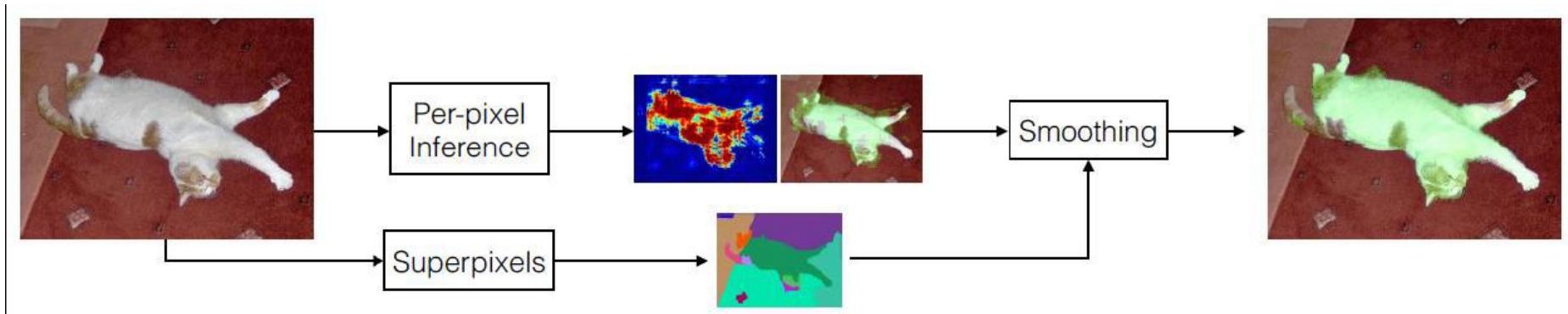
Pinheiro, Weakly Supervised Object Segmentation with Convolutional Neural Networks, 2014

Слабое обучение



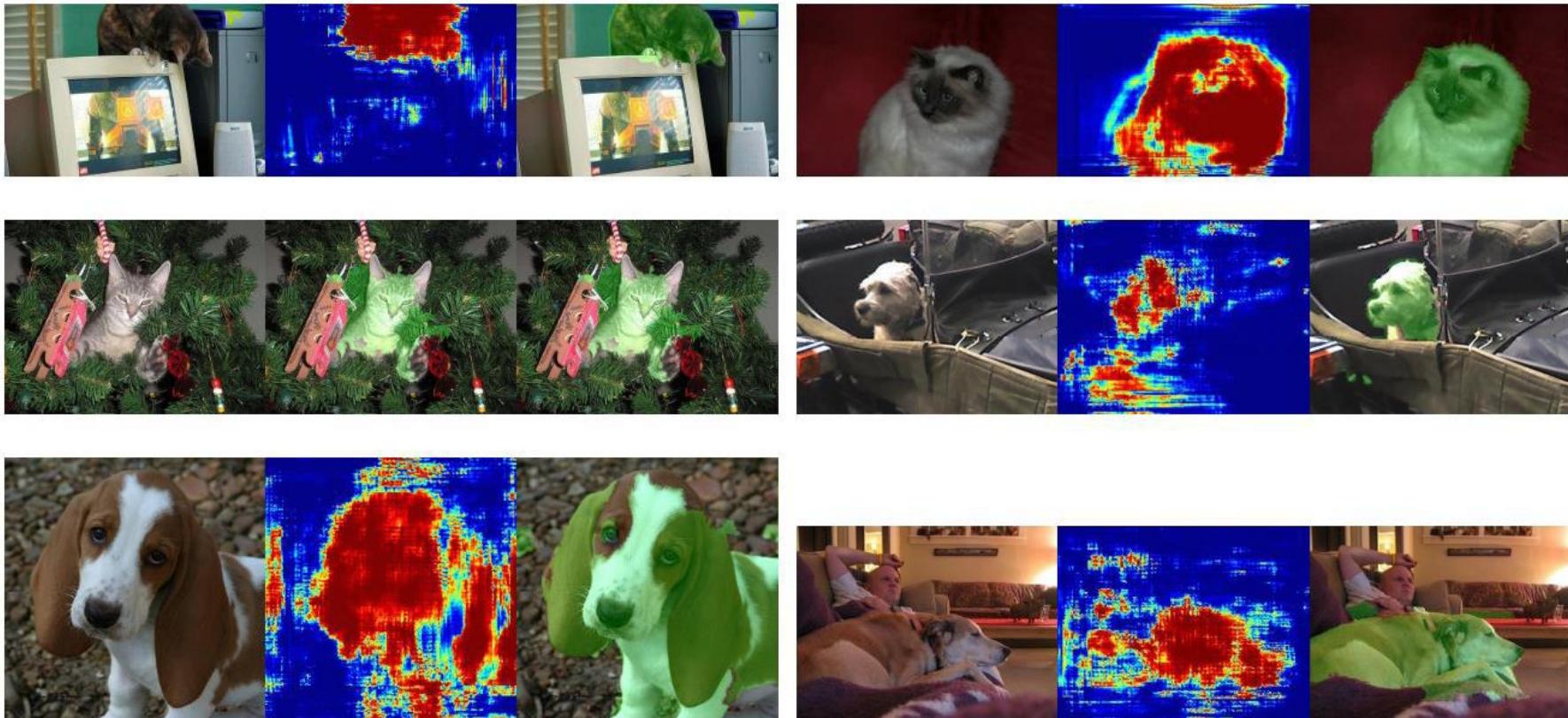
- Фрагменты из изображений, в которых есть объект – положительные примеры
- Фрагменты изображений, в которых нет объекта – отрицательные примеры

Применение полученного классификатора



- Применяем получившийся классификатор попиксельно
- Совмещаем с суперпикселями

Пример результатов



DeepLab

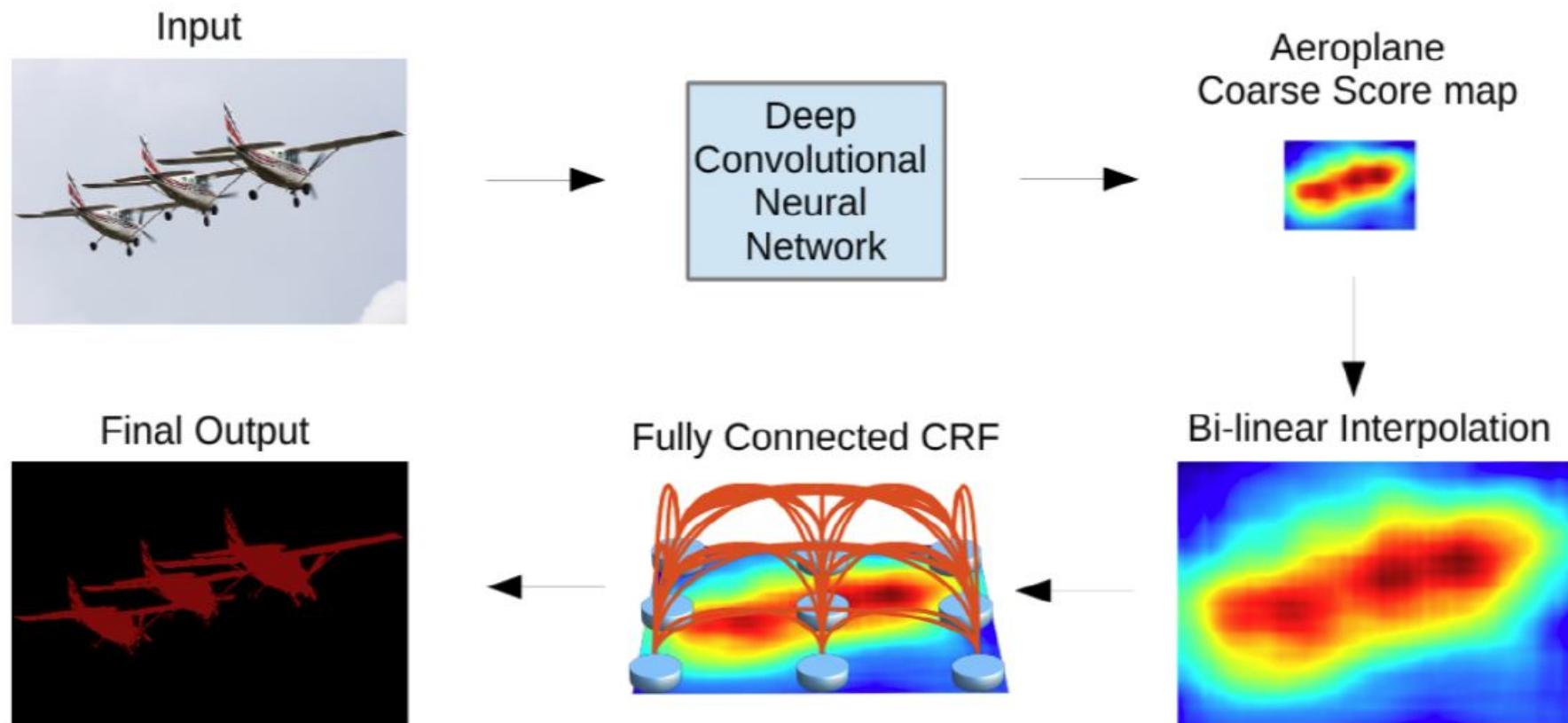


Image credit: Chen et al.



- Базовая VGG-16 (потом заменили на ResNet)
- Превращаем в свёрточную, вытягивая fc слои в 1×1 свёртку
- Прямое использование базовых архитектур уменьшает разрешение в 32 раза
- DeepLab – убираем часть subsampling, получаем $x8$ уменьшение
- Другие подходы используют специальные методы повышения разрешения

Результат нейросети



Границы очень примерные

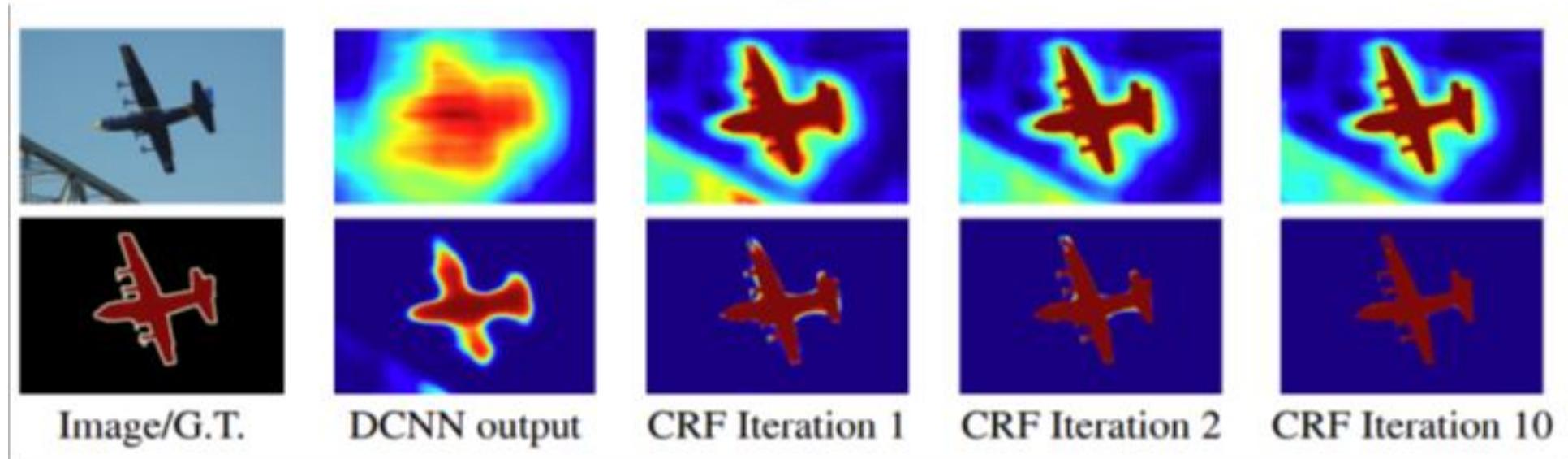


Уточнение через Dense CRF

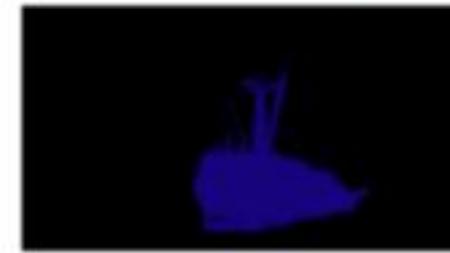


- Гауссовский пространственный парный потенциал + билатеральный парный потенциал
- Гипер-параметры подбираем на валидационной выборке сеточным методом

$$w_1 \exp \left(- \frac{\|p_i - p_j\|^2}{2\sigma_\alpha^2} - \frac{\|I_i - I_j\|^2}{2\sigma_\beta^2} \right) + w_2 \exp \left(- \frac{\|p_i - p_j\|^2}{2\sigma_\gamma^2} \right)$$



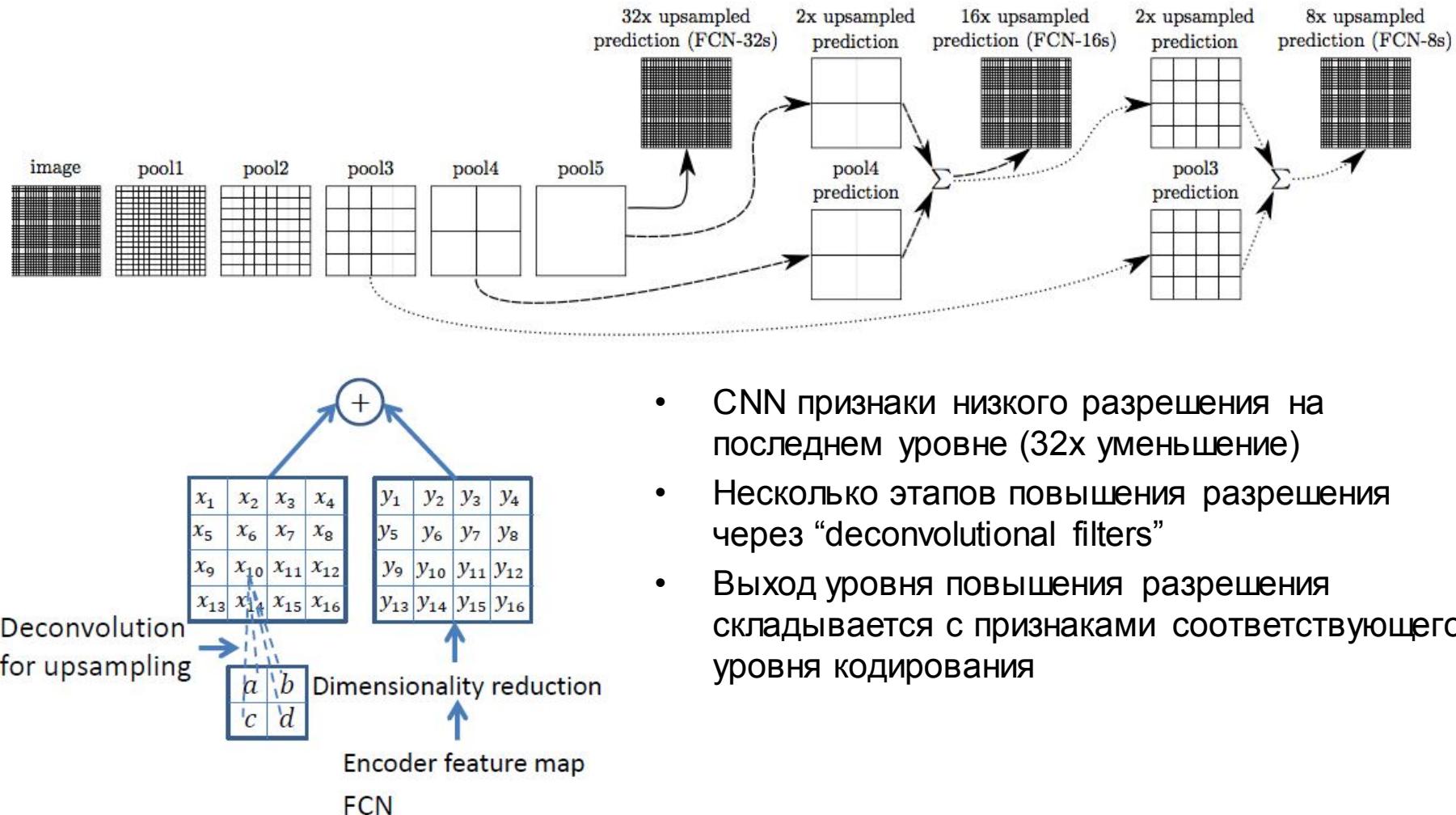
Результаты



Raw score maps

After dense CRF

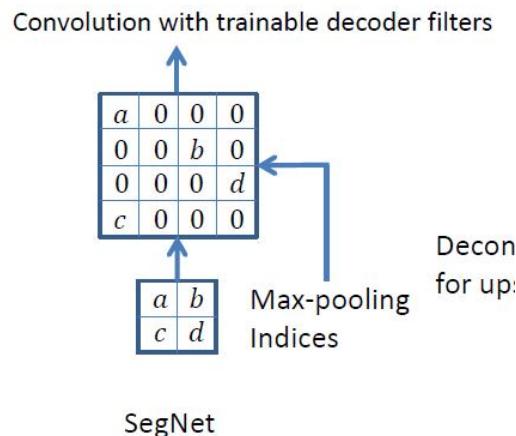
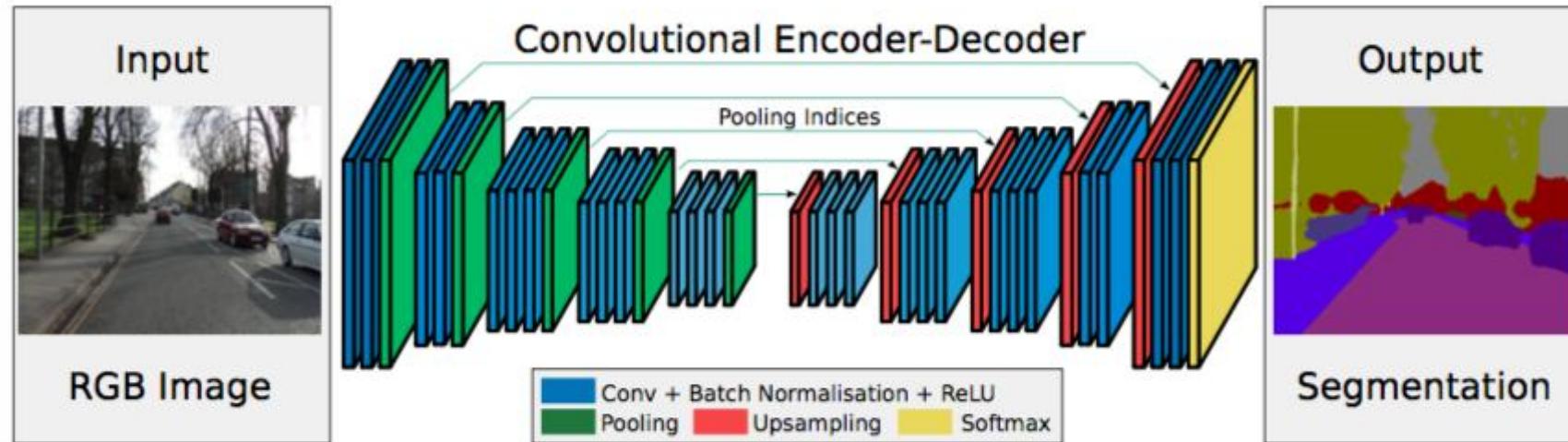
Fully convolutional networks



- CNN признаки низкого разрешения на последнем уровне (32x уменьшение)
- Несколько этапов повышения разрешения через “deconvolutional filters”
- Выход уровня повышения разрешения складывается с признаками соответствующего уровня кодирования

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. CoRR, abs/1411.4038, 2014.

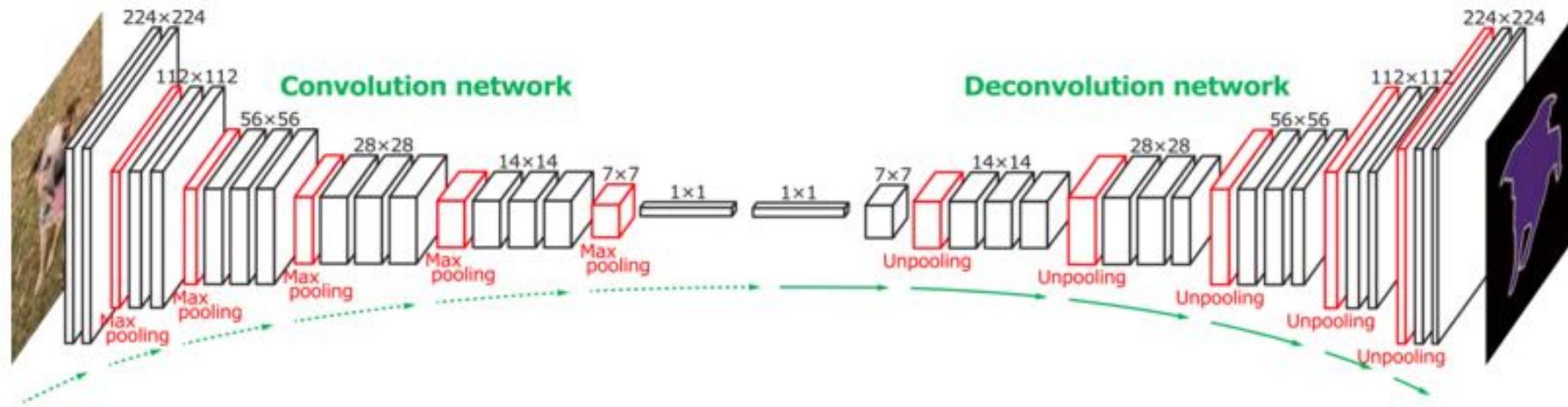
Encoder-Decoder



- Меняем схему «повышения разрешения»
- Сохраняем индексы каждого max-pooling слоя
- Про повышении разрешения делаем так:
 - Копируем значения из выхода max-pooling слоя с учётом запомненный индексов
 - Применяем обученные свёртки для сглаживания
- Делаем в несколько этапов до исходного разрешения

Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. CoRR, abs/1511.00561, 2015.

Deconvolutional network

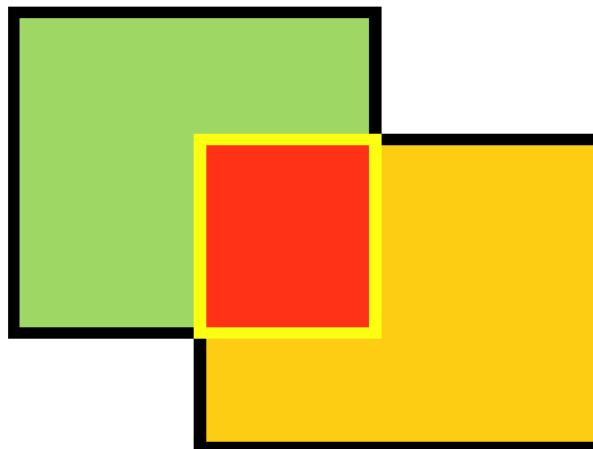


- Доведение идеи encoder-decoder до конца
 - По всему изображению строим вектор признаков 1×1
 - Используя сохранённые max-pooling индексы повышаем разрешение до исходного, предсказывая карту разметки

Оценка точности



- Intersection over Union



$$\text{segmentation accuracy} = \frac{\text{true positives}}{\text{true positives} + \text{false positives} + \text{false negatives}}$$

Развитие методов на PASCAL VOC



- До CNN (<50%)
- CNN (60---64%)
- CNN + CRF (67%)
- Data Augmentation (71%)
 - Предобучение на других данных (COCO или Pascal)
 - Weakly supervision с помощью bbox
- Совместное обучение параметров CRF и CNN (74%)
- Обучение попарной совместимости меток (77%)

Наборы данных



- MSRC
 - SiftFlow
 - Stanford
 - LabelMe
 - NYU
 - Sun RGBD
 - Pascal VOC
 - Microsoft COCO
 - KITTI
 - CamVid
 - Cityscapes
- Classic but might not large enough for data-hungry models
- RGBD-benchmark, large, indoor
Good for CNNs
- Object labeling, large, object seg
Good for CNNs, good for multi-task
- Autonomous driving, video, scene
Enough for fine-tuning CNNs
Good for multi-task
Good for taking two courses

Резюме лекции



- Важные факторы при сегментации
 - Локальная классификация
 - Локальный контекст
 - Глобальный контекст
- Основные инструменты
 - Пересегментация (суперпиксельная сегментация)
 - Графические модели (CRF, Dense CRF)
 - Нейросети
- Текущее развитие
 - Интеграция нейросетей и графических моделей, совместное обучение