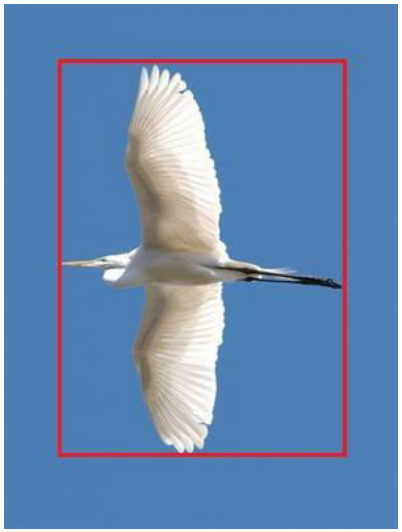
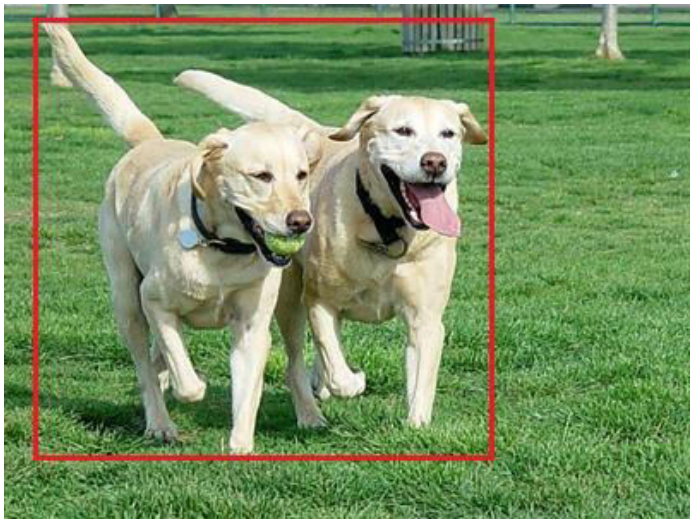
Обнаружение выделяющихся инородных объектов на базе сверточной нейронной сети и векторов Фишера

Асп. 2 г.о. Саушкин Роман

Выделяющийся инородный объект

- Выделяющийся инородный (salient) объект – ярко выраженный объект, по своей структуре отличающийся от других объектов сцены. Другими словами, это такой объект, на котором в первую очередь фиксируется взгляд и привлекается внимание человека.





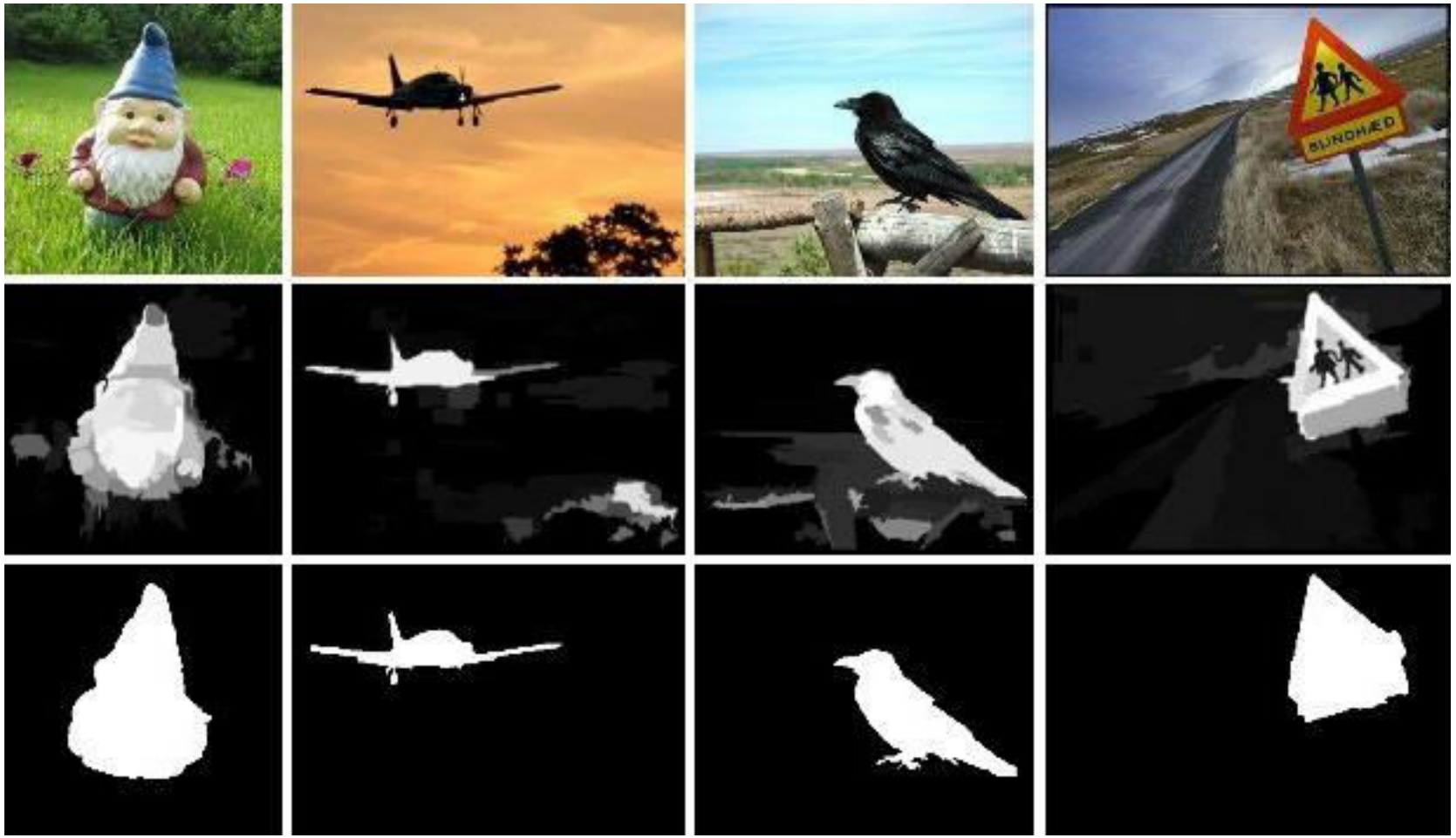
Где впервые возникла задача обнаружения выделяющихся инородных объектов

- Задача предсказания областей фиксации взгляда человека в целях исследования механизмов работы зрительной системы человека
- Задача определения **ROI** (**Region Of Interest** — регион интересов — интересующая область изображения)

Области применения сегодня

- Кадрирование изображений (image cropping)
- Ретаргетинг изображений (image retargeting)
- Описание содержания изображения (image summarization)
- Генерация эскизов (thumbnail generation)
- Слежение за объектом (object tracking)
- Классификация изображений (image classification)
- Идентификация личности (person re-identification)

Карта выделяемости (saliency map)

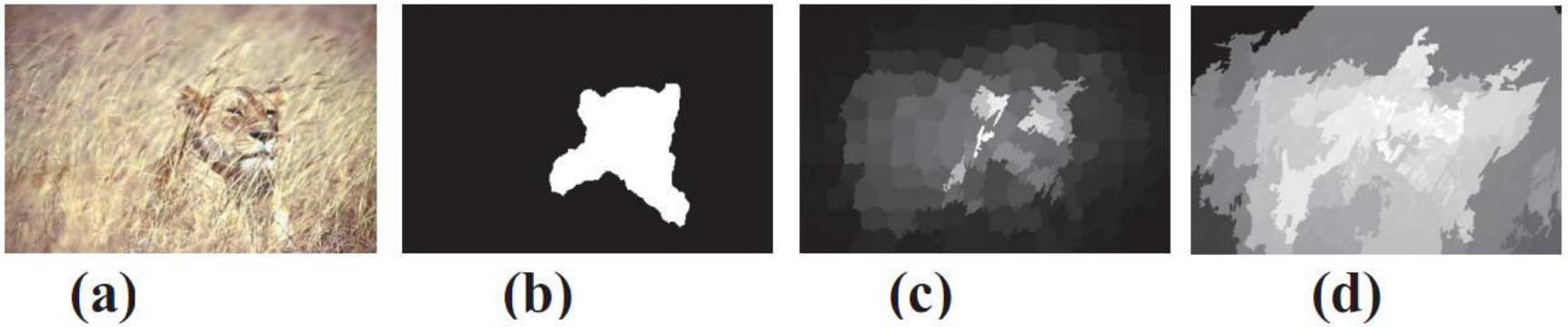


Контрастность

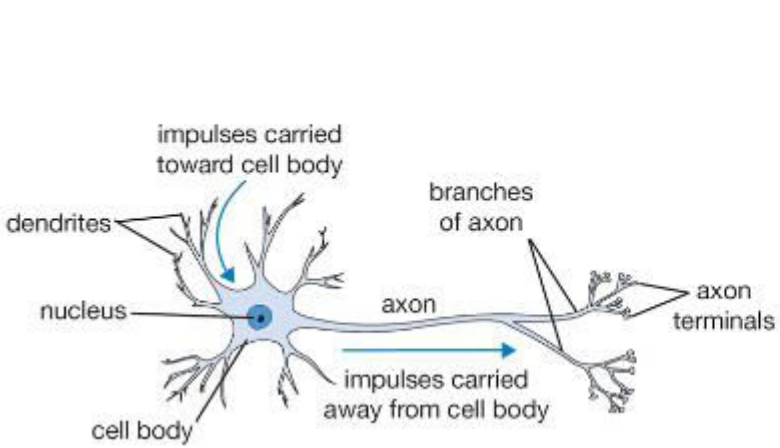
- Контрастность – наиболее существенный фактор для зрительного внимания (привлекает внимание человека на низком уровне)
- Традиционно контрастность описывается с помощью признаков, разработанных самим человеком. Они основываются на сравнении таких низкоуровневых характеристик, как цвет, яркость, гистограммы, характеризующие текстуру, и т.п.

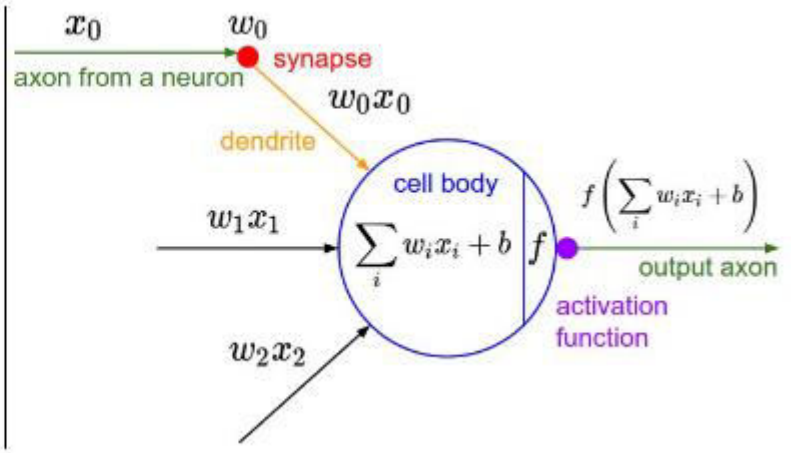
Недостаток:

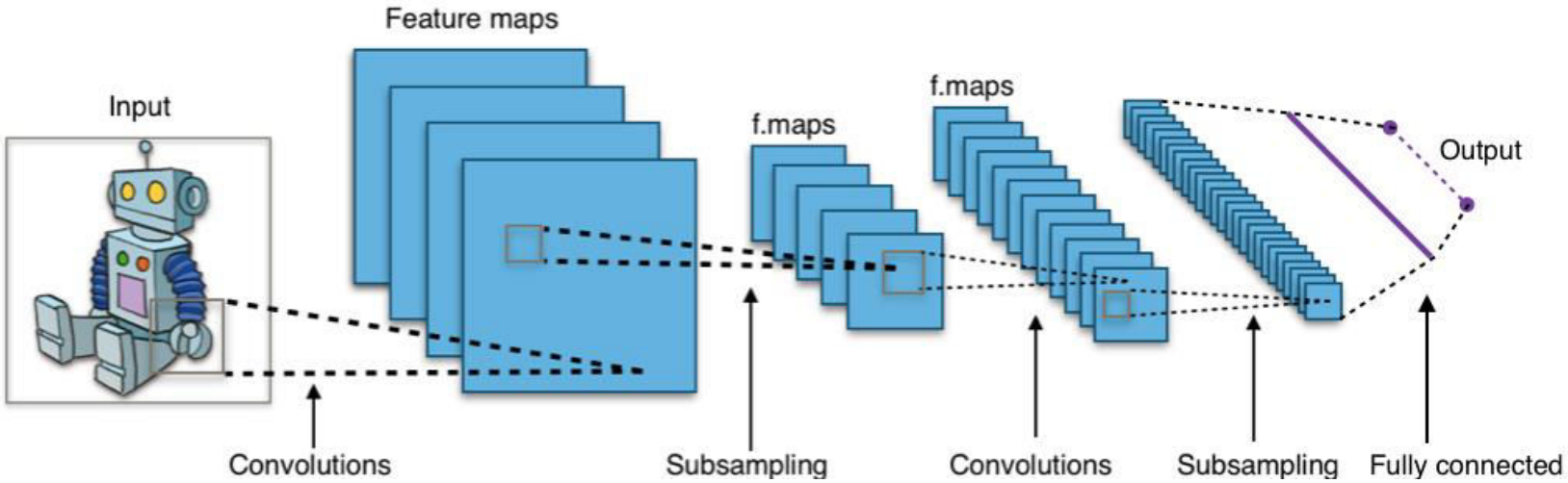
- Низкоуровневые признаки не способны эффективно «улавливать» семантический контекст, скрытый в изображении



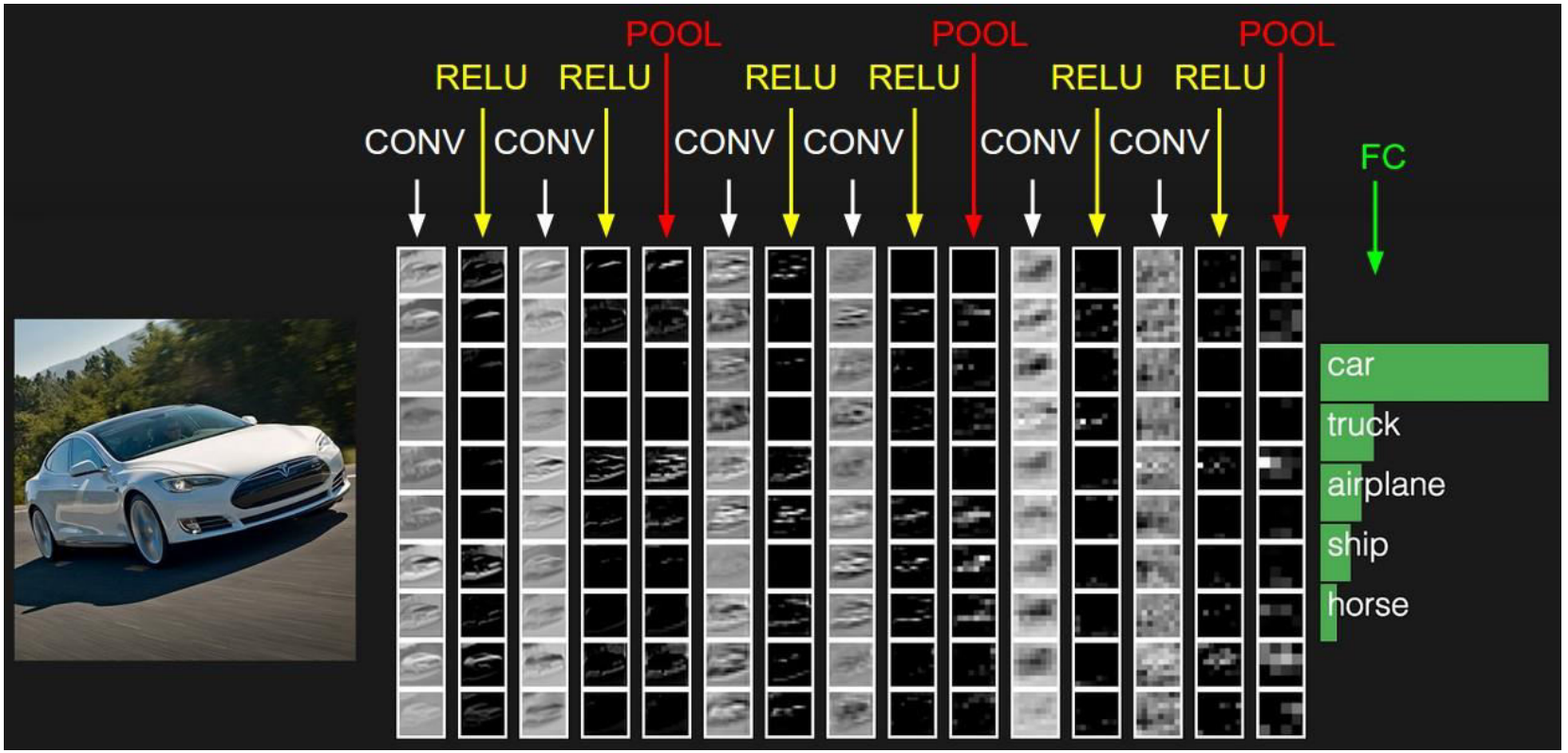
Сверточная нейронная сеть







Сверточная нейронная сеть (CNN)

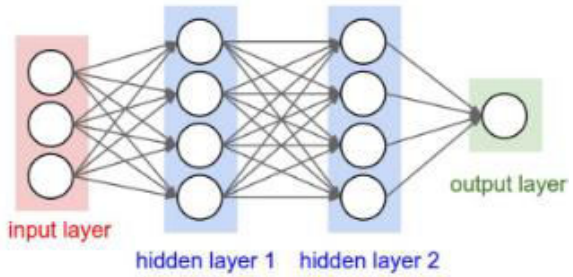


Sigmoid $\sigma(x) = 1/(1+e^{-x})$ Tanh. $\tanh(x) = 2\sigma(2x) - 1$.

ReLU. $f(x) = \max(0, x)$.

Комбинация CNN- и низкоуровневых признаков

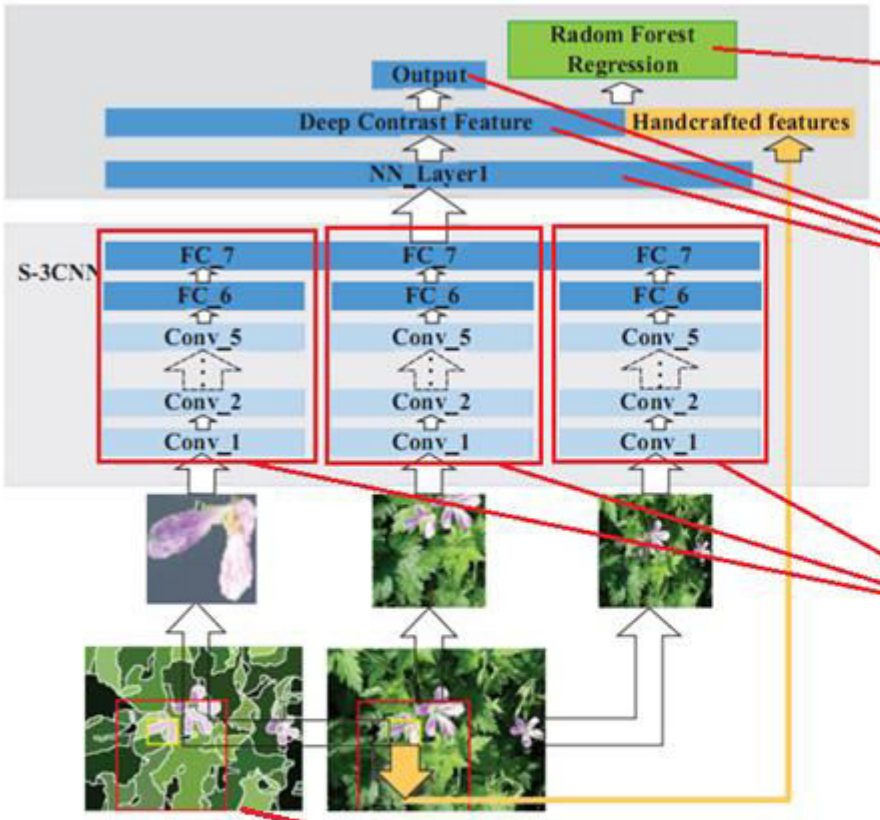
Features
Difference between Average RGB Values
χ^2 distance between RGB Histograms
Difference between Average LAB Values
χ^2 distance between LAB Histograms
Difference between Average HSV Values
χ^2 distance between HSV Histograms
χ^2 distance b.w. Max response LM Histograms
χ^2 distance between LBP Histograms



CNN-признаки

Низкоуровневые признаки

Общая схема алгоритма



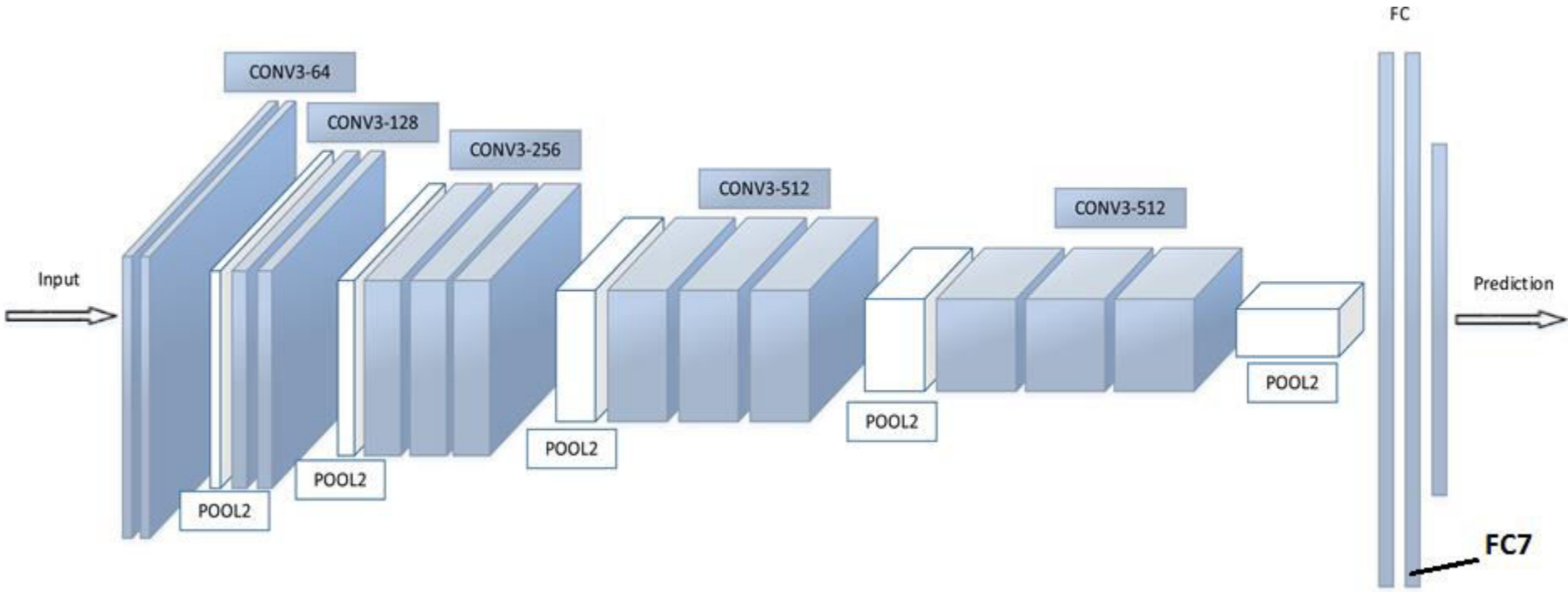
Случайный лес
- обучается на комбинации низкоуровневых признаков и признаков, полученных из нейронной сети

Нейронная сеть с двумя скрытыми слоями
- обучается на наборе помеченных сегментов обучающей выборки
- используется в качестве регрессора с одной стороны и для дальнейшего извлечения признаков с другой

3 x VGG-M
- обучены на ImageNet
- адаптированы для задачи обнаружения объектов (object detection)
- используются для извлечения признаков сегментов

Efficient Graph-Based Image Segmentation (Pedro F. Felzenszwalb, Daniel P. Huttenlocher)

Архитектура VGG-M



- 5 сверточных слоев
- 3 полностью соединенных слоя
- Входной слой принимает изображения размером 224x224
- Признаки извлекаются из предпоследнего полностью соединенного слоя, **FC7**
- Выход из **FC7** – вектор размерности 4096

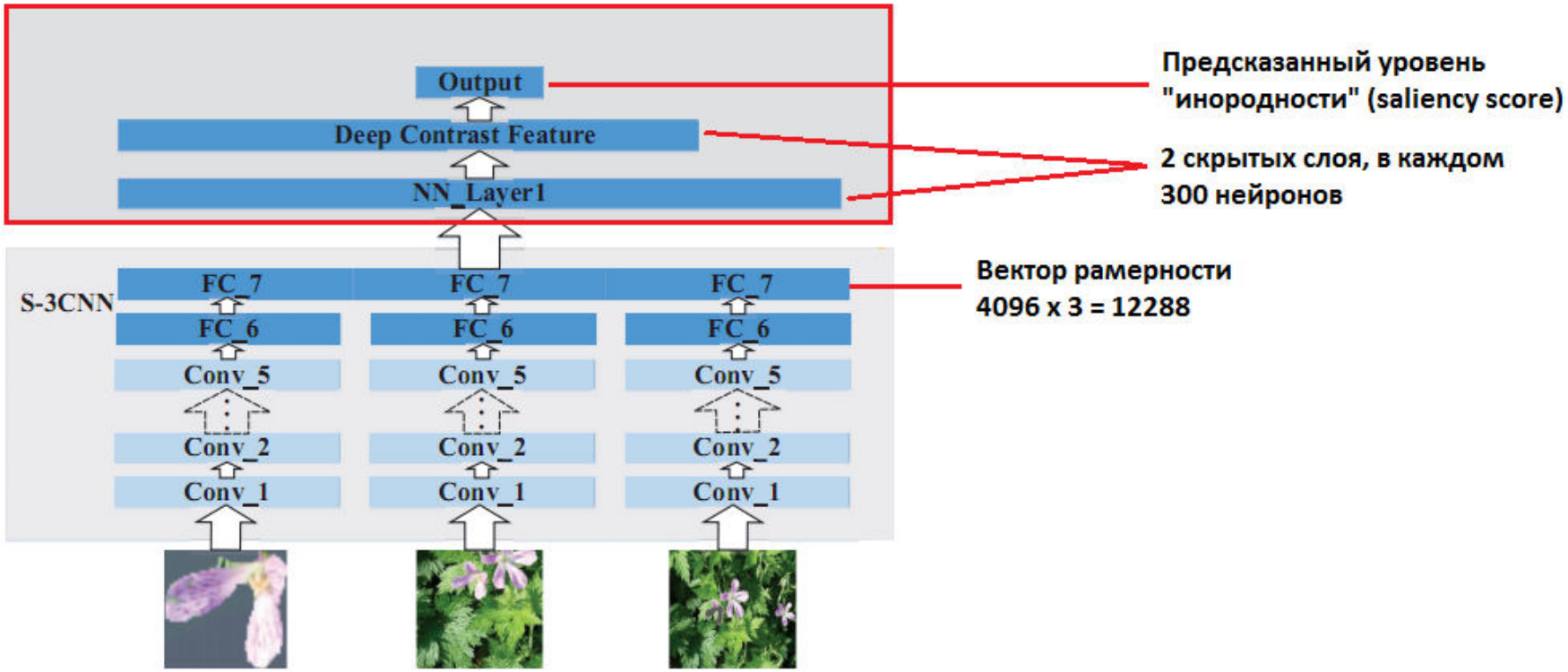
Предварительная обработка сегментов

А. Сегмент помещается в ограничивающий прямоугольник,и его размер изменяется до 224x224. Область вне сегмента заполняется пикселями из среднего изображения

Б. Сегмент и непосредственно граничащие с ним сегменты помещаются в ограничивающий прямоугольник, и его размер изменяется до 224x224

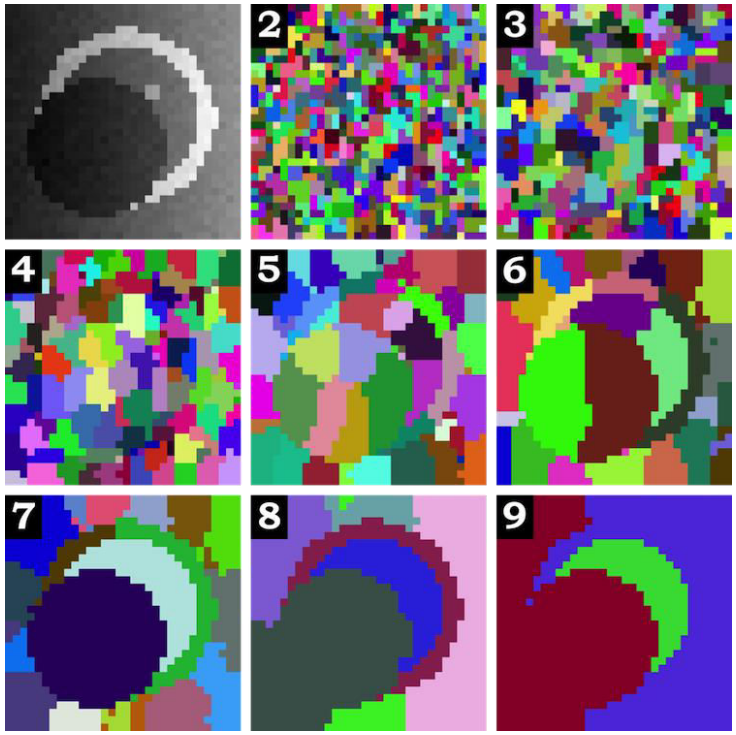
В. Размер всего изображения изменяется до 224x224, при этом область, соответствующая сегменту, заполняется пикселями из среднего изображения

Обучение нейронной сети



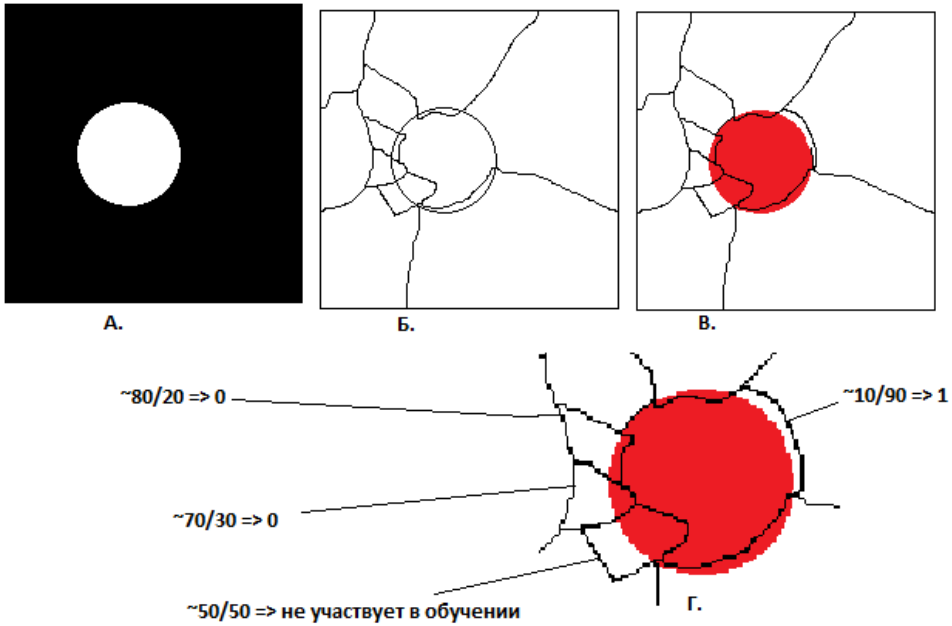
Обучение нейронной сети

- Перед обучением проводится сегментирование изображения на 15 уровнях. На каждом уровне устанавливается максимальное количество сегментов изображения. На первом уровне это 300 сегментов. На последнем – 20 сегментов. Сегменты со всех уровней участвуют в обучении единственной нейронной сети.



Обучение нейронной сети

- Каждому сегменту присваивается метка: 0 – сегмент не является инородным, 1 – сегмент является инородным.
- Решение о том, принимает ли участие сегмент в обучении, принимается по следующему правилу:
  - Если сегмент на >= 70% состоит из истинно инородных пикселей, то он участвует в обучении и ему присваивается метка 1.
  - Если сегмент на >= 70% состоит из истинно **не**инородных пикселей, то он участвует в обучении и ему присваивается метка 0.
  - Иначе сегмент не участвует в обучении
- Данную часть метода назовем **MDF**(Multiscale Deep Features).



Объединение карт выделяемости

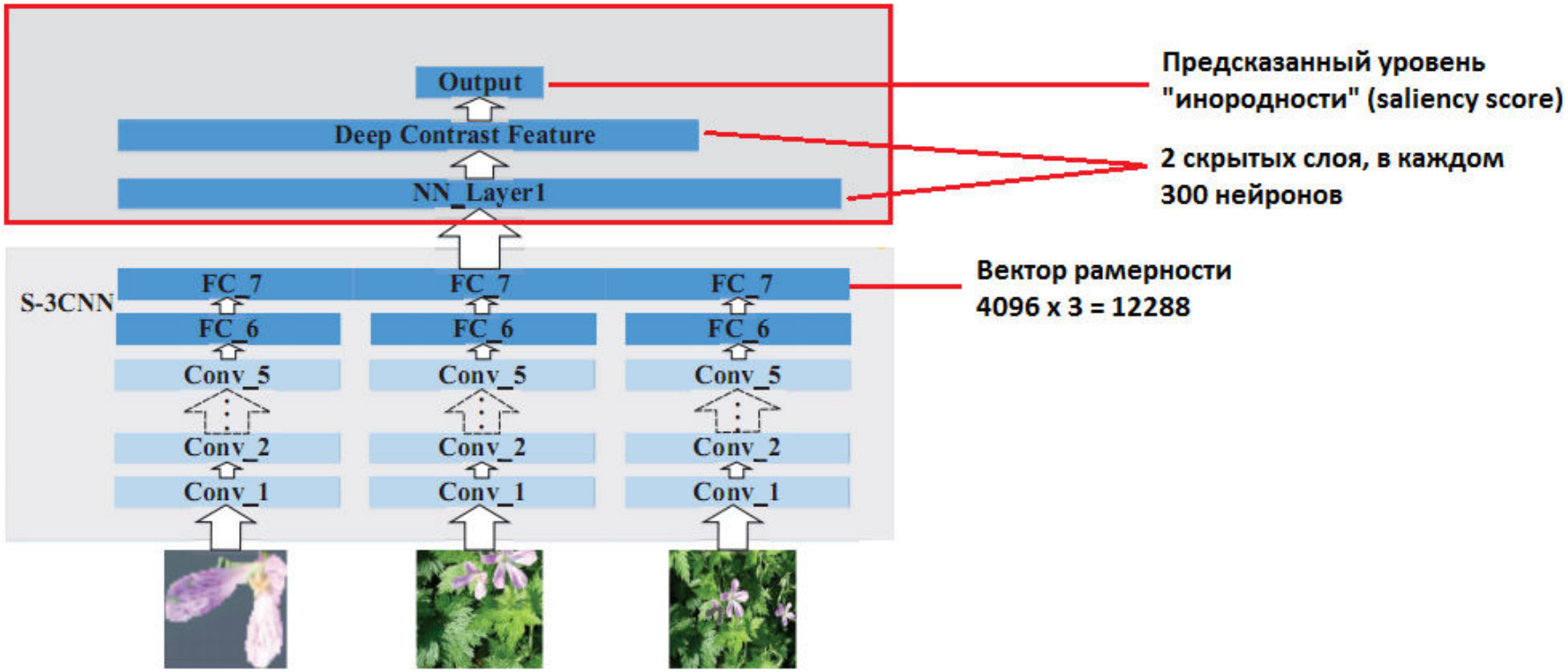
- Задача:
  - Дано: M уровней сегментации изображений
  - Получено: M карт выделяемости для каждого уровня сегментации
  - Требуется: объединить M карт выделяемости в одну суммарную
- Решение:
  - Взять линейную комбинацию карт выделяемости с весами α_k.
  - Найти веса с помощью метода наименьших квадратов, используя валидационное множество.

$$A = \sum_{k=1}^{M} \alpha_k A^{(k)}$$

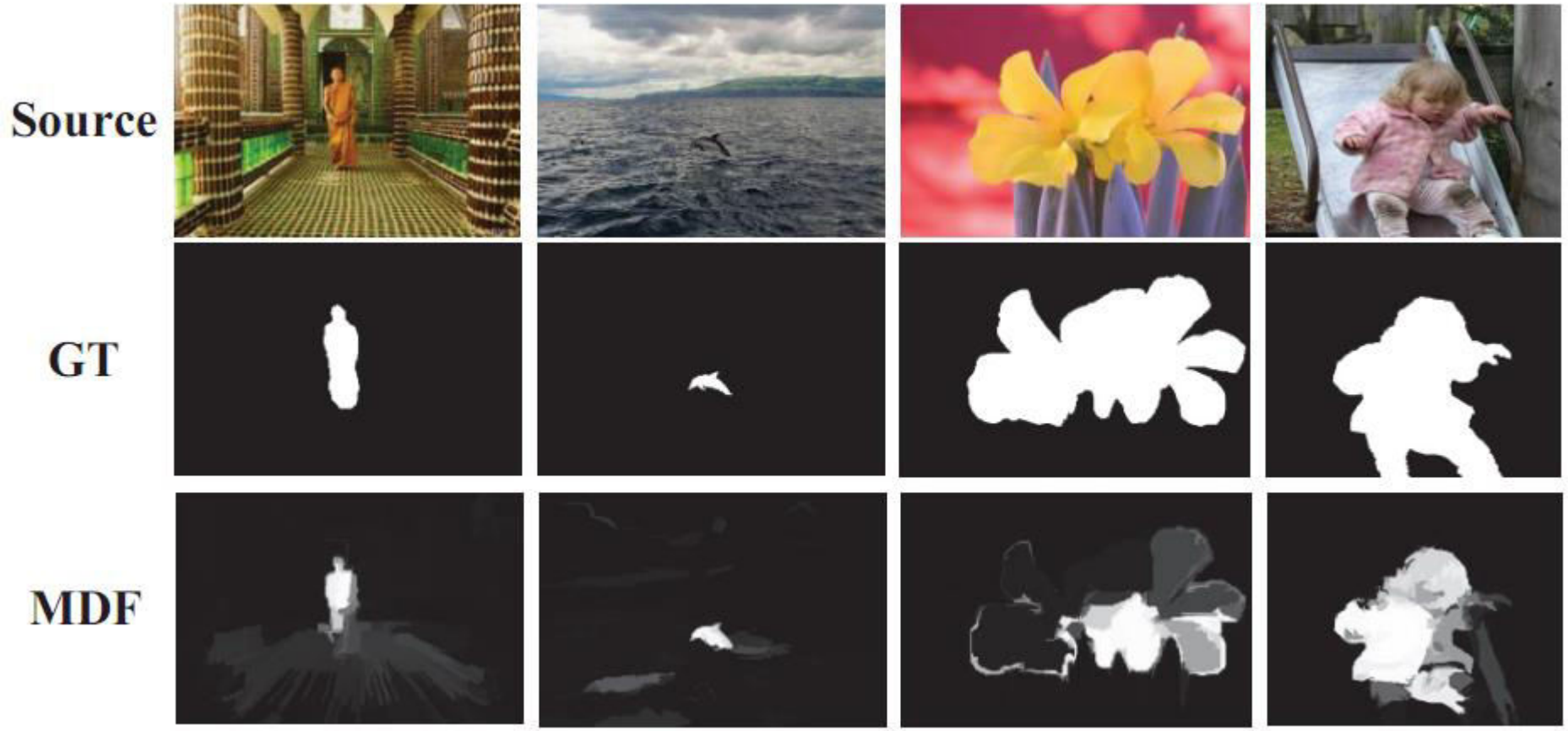
$$\text{s.t. } \{\alpha_k\}_{k=1}^{M} = \underset{\alpha_1,\alpha_2,\ldots,\alpha_M}{\operatorname{argmin}} \sum_{i \in I_v} \left\| A_i - \sum_k \alpha_k A_i^{(k)} \right\|_F^2$$

I_v- множество индексов изображений валидационного множества

Промежуточные результаты

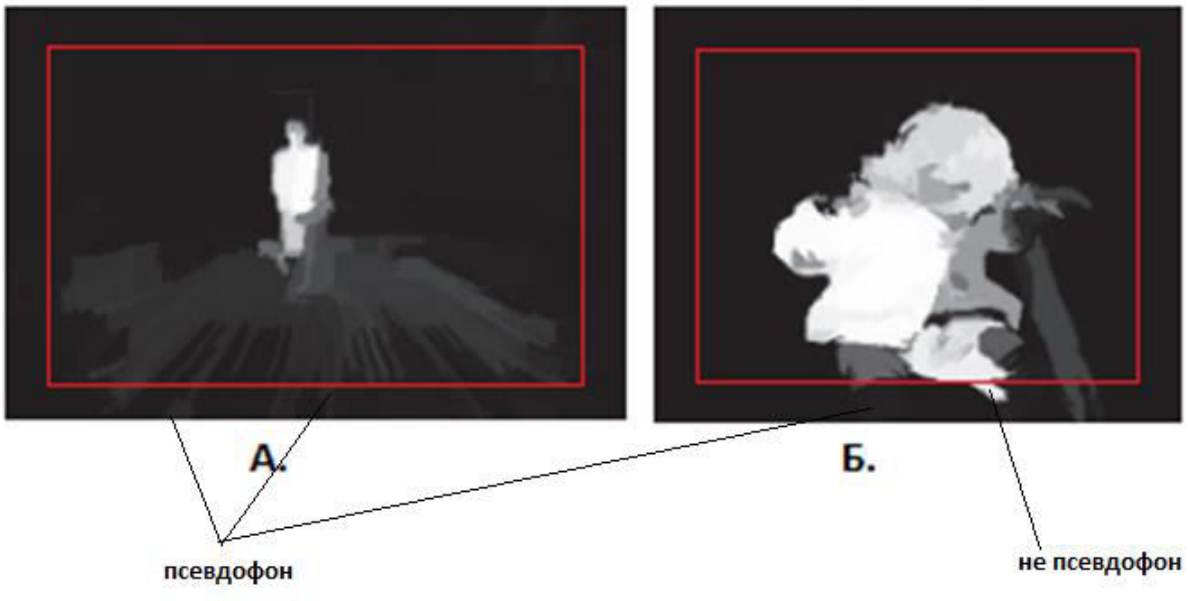


Промежуточные результаты



Интеграция с низкоуровневыми признаками

- Для каждого изображения определяется начальная карта выделяемости Sm_{init} с помощью MDF.
- Определяется псевдофоновая область B как множество пикселей, отстоящие не более чем на 30 пикселей от границы изображения и которые имеют $Sm_{init} < 0.1$

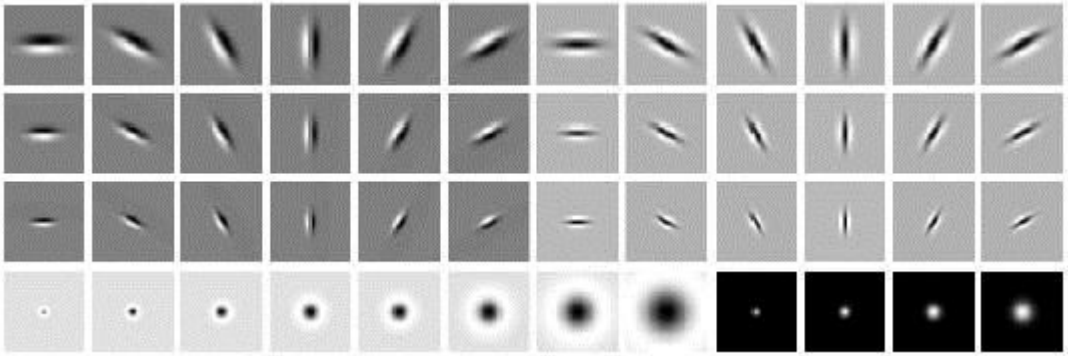


Интеграция с низкоуровневыми признаками

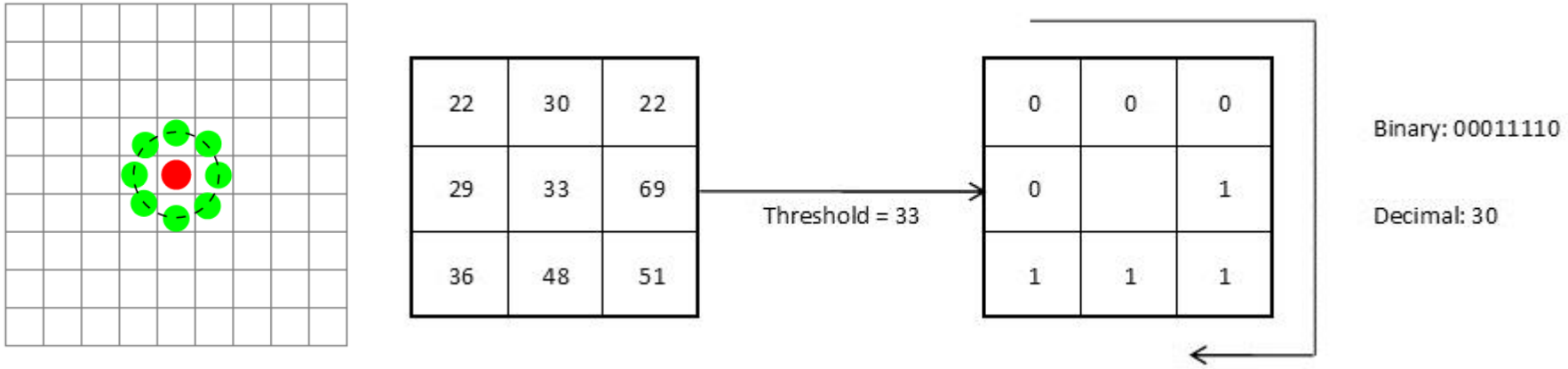
- Для всего изображения, области B и всех сегментов на всех уровнях сегментации вычисляются признаки, описывающие цвет (RGB, HSV, LAB – гистограммы и их средние значения) и текстуру (LM - фильтры, LBP - признаки).
- По полученным признакам вычисляются низкоуровневые контрастные признаки.

LM – фильтры, LBP

- LM(Leung-Malik) - фильтры



- LBP – Local Binary Patterns

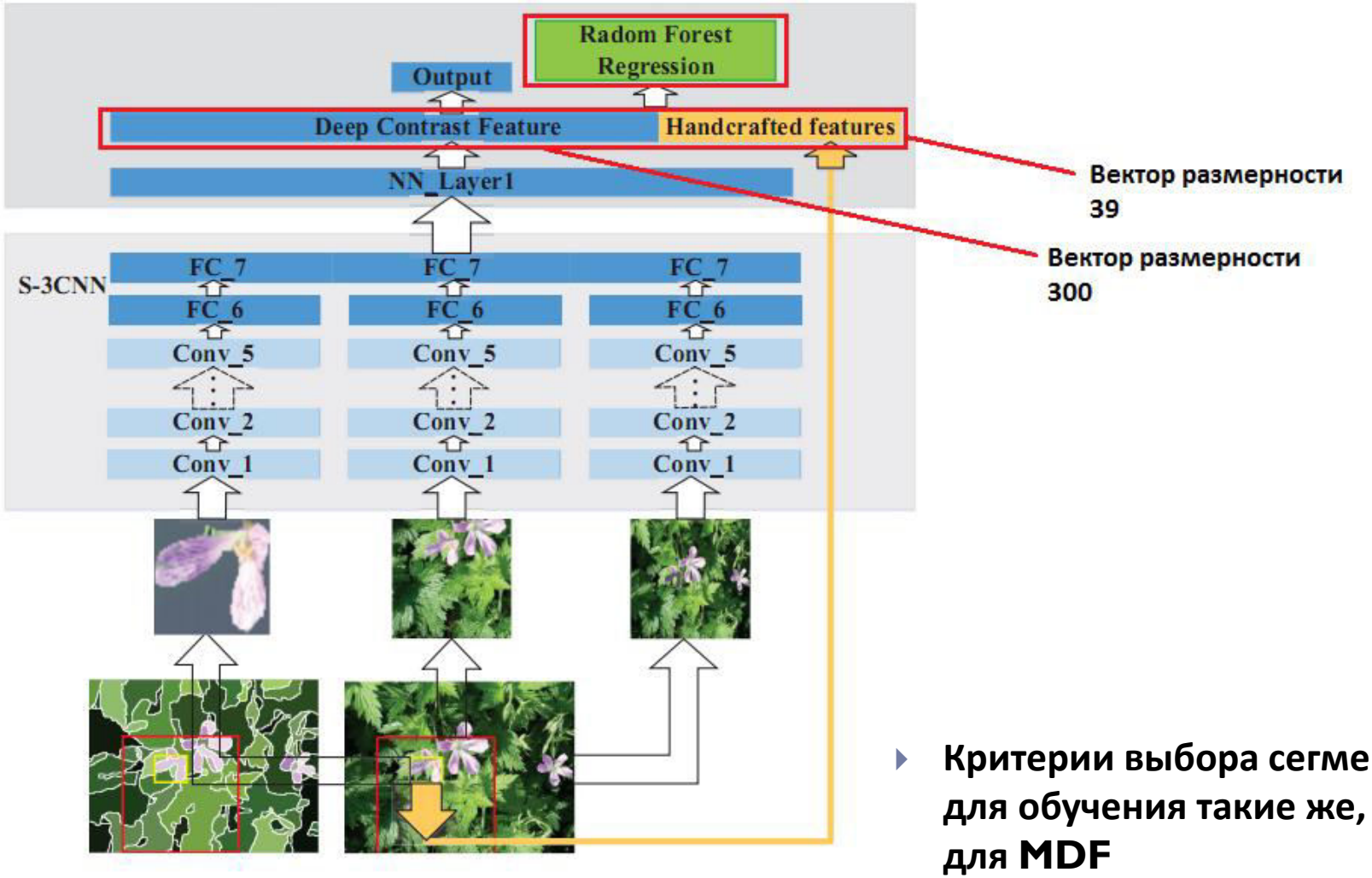


Низкоуровневые контрастные признаки

Contrast Descriptors (Color and Texture)				Segment Properties			
Notation	Features	Definition	Dim	Notation	Features	Definition	Dim
$c_1 \sim c_6$	Difference between Average RGB Values	$\lvert R^{rgb} - B^{rgb} \rvert$, $\lvert R^{rgb} - I^{rgb} \rvert$	6	$s_1 \sim s_3$	Variances of RGB values	var_R^r, var_R^g, var_R^b	3
$c_7 \sim c_8$	χ^2 distance between RGB Histograms	$\chi^2(h_{rgb}^R, h_{rgb}^B)$, $\chi^2(h_{rgb}^R, h_{rgb}^I)$	2	$s_4 \sim s_6$	Variances of LAB values	var_R^l, var_R^a, var_R^b	3
$c_9 \sim c_{14}$	Difference between Average LAB Values	$\lvert R^{lab} - B^{lab} \rvert$, $\lvert R^{lab} - I^{lab} \rvert$	6	$s_7 \sim s_9$	Variance of HSV values	var_R^h, var_R^s, var_R^v	3
$c_{15} \sim c_{16}$	χ^2 distance between LAB Histograms	$\chi^2(h_{lab}^R, h_{lab}^B)$, $\chi^2(h_{lab}^R, h_{lab}^I)$	2	s_{10}	Normalized perimeter	$Perimeter(R)$	1
$c_{17} \sim c_{22}$	Difference between Average HSV Values	$\lvert R^{hsv} - B^{hsv} \rvert$, $\lvert R^{hsv} - I^{hsv} \rvert$	6	s_{11}	Normalized area	$Area(R)$	1
$c_{23} \sim c_{24}$	χ^2 distance between HSV Histograms	$\chi^2(h_{hsv}^R, h_{hsv}^B)$,$\chi^2(h_{hsv}^R, h_{hsv}^I)$	2				
$c_{25} \sim c_{26}$	χ^2 distance b.w. Max response LM Histograms	$\chi^2(h_{LM}^R, h_{LM}^B)$, $\chi^2(h_{LM}^R, h_{LM}^I)$	2				
$c_{27} \sim c_{28}$	χ^2 distance between LBP Histograms	$\chi^2(h_{LBP}^R, h_{LBP}^B)$, $\chi^2(h_{LBP}^R, h_{LBP}^I)$	2				

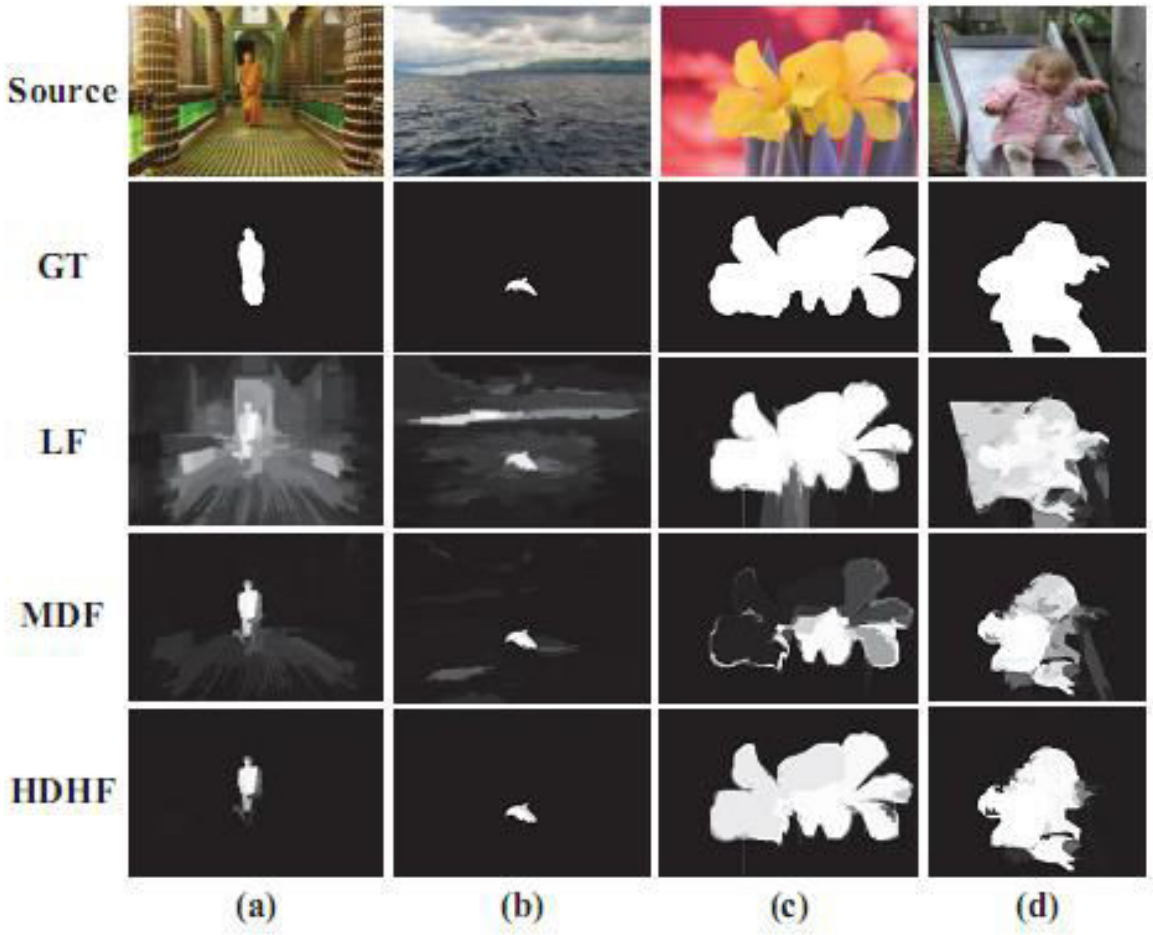
- R – сегмент
- B – псевдофон
- I – целое изображение

Обучение случайного дерева



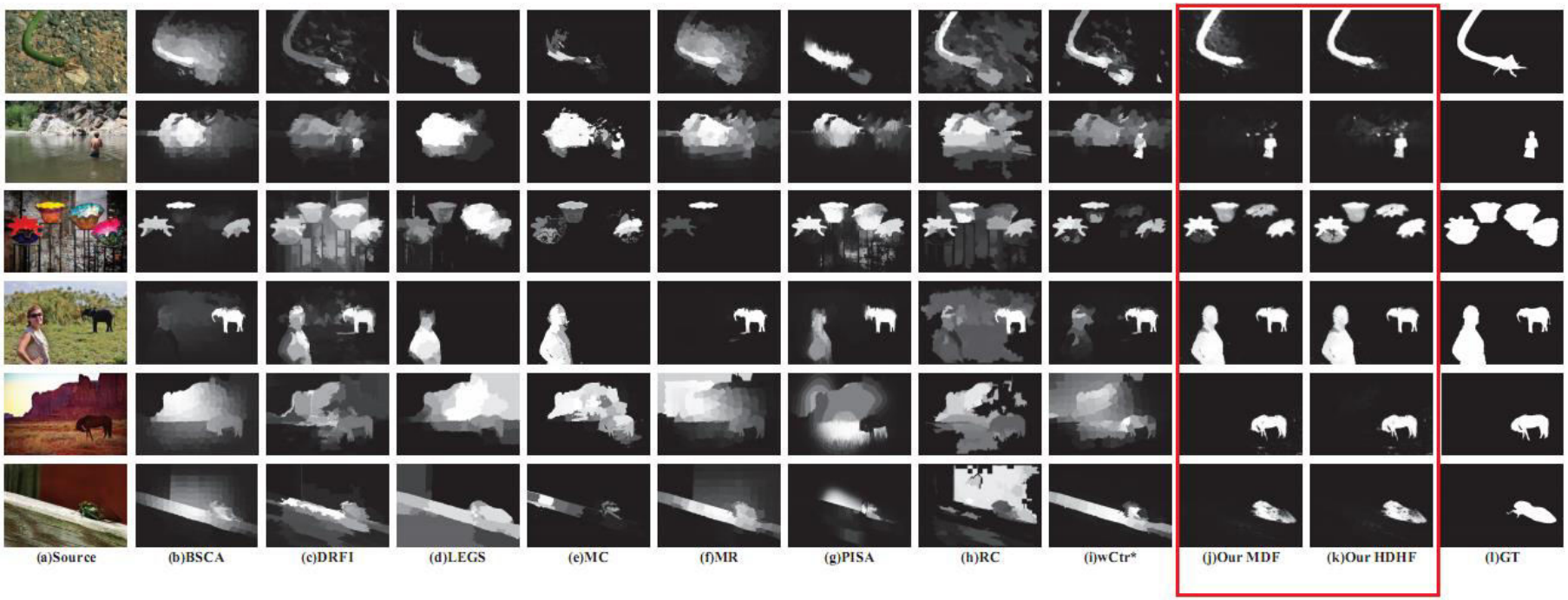
- **Критерии выбора сегментов для обучения такие же, как для MDF**

Сравнение результатов



- GT (ground true)– истинные карты выделяемости
- LF (low-level features) – карты выделяемости, полученные методом обучения случайного дерева только на низкоуровневых признаках
- MDF (multiscale deep features) – карты выделяемости, полученные методом обучения только нейронной сети (без использования низкоуровневых признаков)
- HDHF (hybrid deep and handcrafted feature) – карты выделяемости, полученные методом обучения случайного дерева на признаках из нейронной сети и низкоуровневых признаках

Сравнение с другими методами



Кодирование изображения вектором Фишера

- **Вектор Фишера** – это представление изображения, полученное в результате пулинга локальных дескрипторов изображения.

Обучающий набор изображений

Локальные дескрипторы изображения обучающего набора - векторы признаков размерности D

Локальные дескрипторы изображения - векторы признаков размерности D

$$I = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$$

$$q_{ik} = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_k)^T \Sigma_k^{-1}(\mathbf{x}_i - \mu_k)\right]}{\sum_{t=1}^{K} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \mu_t)^T \Sigma_k^{-1}(\mathbf{x}_i - \mu_t)\right]}$$

$$u_{jk} = \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^{N} q_{ik} \frac{x_{ji} - \mu_{jk}}{\sigma_{jk}},$$

$$v_{jk} = \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^{N} q_{ik} \left[\left(\frac{x_{ji} - \mu_{jk}}{\sigma_{jk}} \right)^2 - 1 \right]$$

k = I,…,K

j = I,…, D

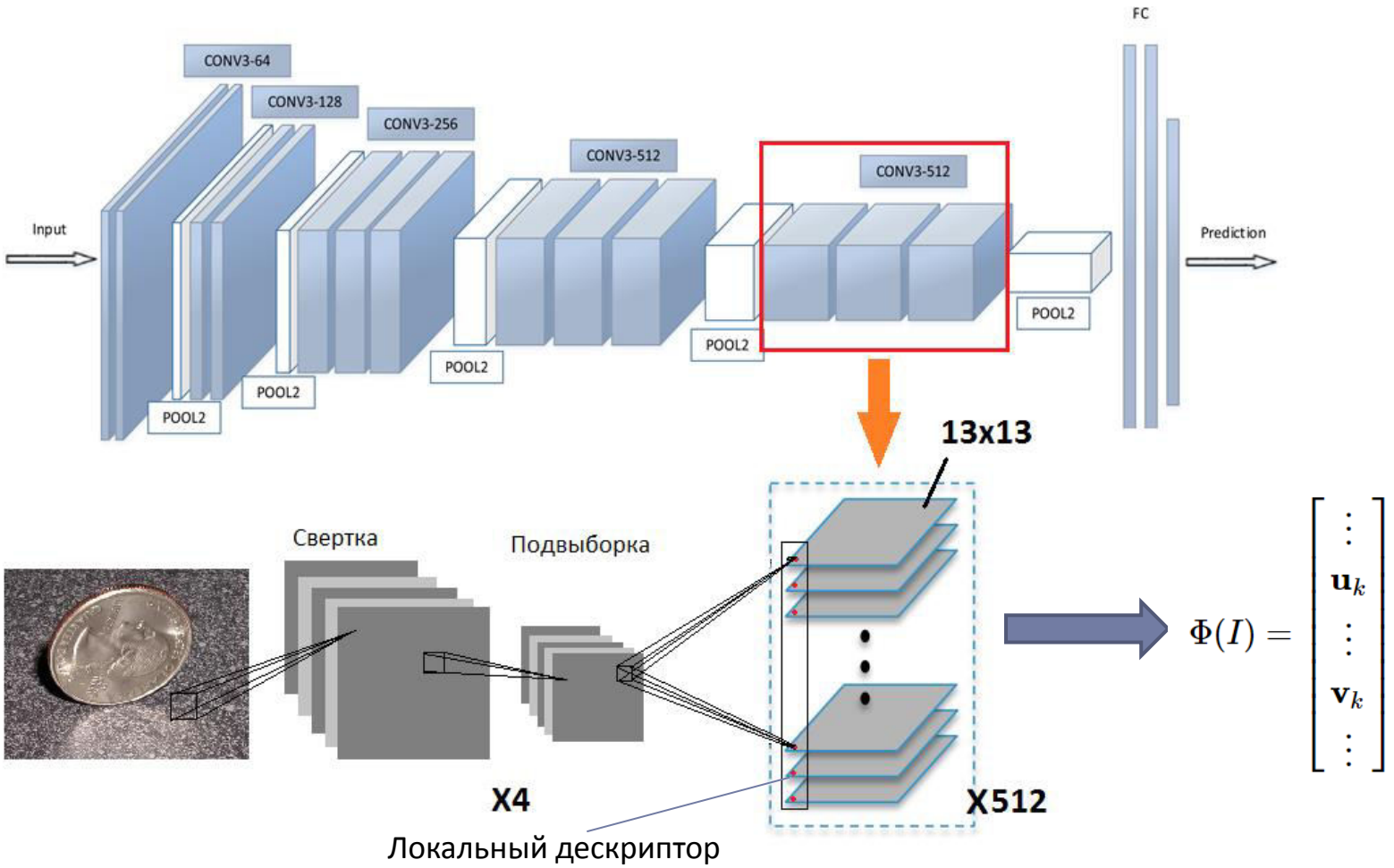
Модель смеси гауссовых распределений(GMM), описывающая распределение локальных дескрипторов.

Параметры модели:

$$\Theta = (\mu_k, \Sigma_k, \pi_k : k = 1, \ldots, K)$$

$$\Phi(I) = \begin{bmatrix} \vdots \\ \mathbf{u}_k \\ \vdots \\ \mathbf{v}_k \\ \vdots \end{bmatrix}$$ - вектор Фишера

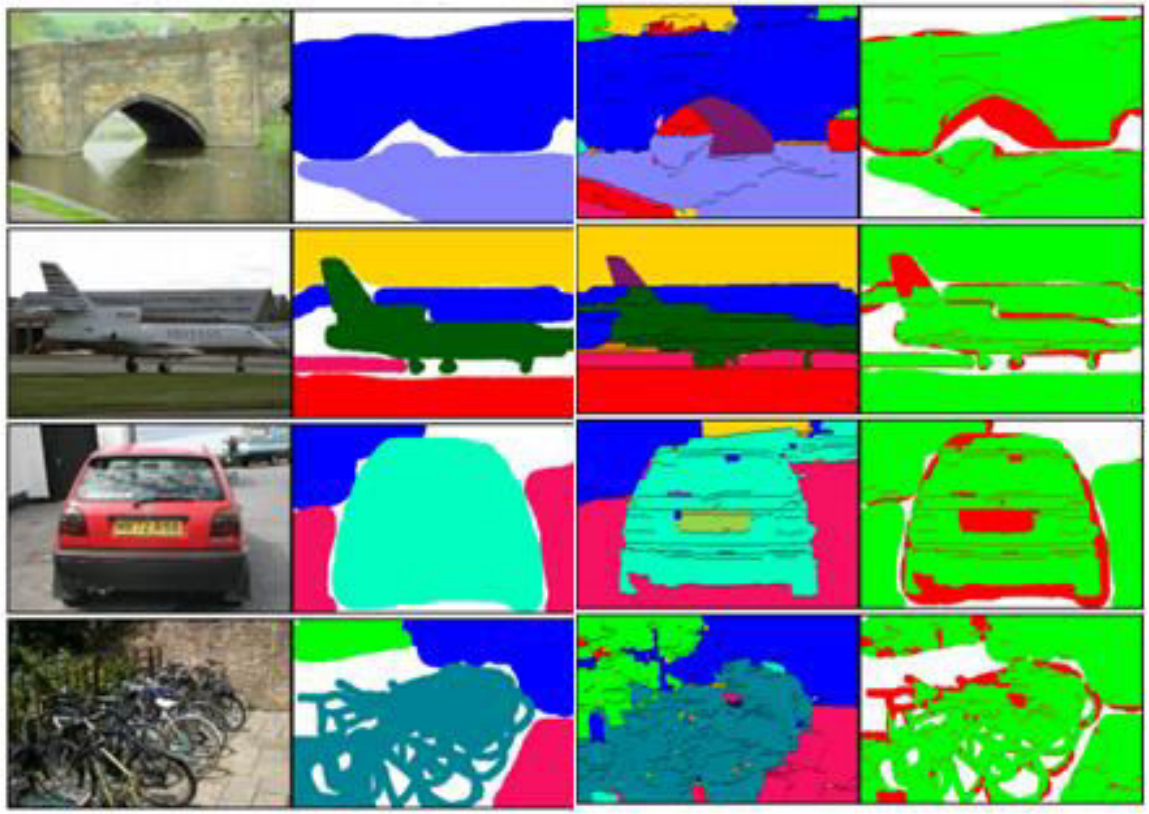
Неупорядоченное кодирование вектором Фишера локальных признаков, полученных с помощью сверточной нейронной сети



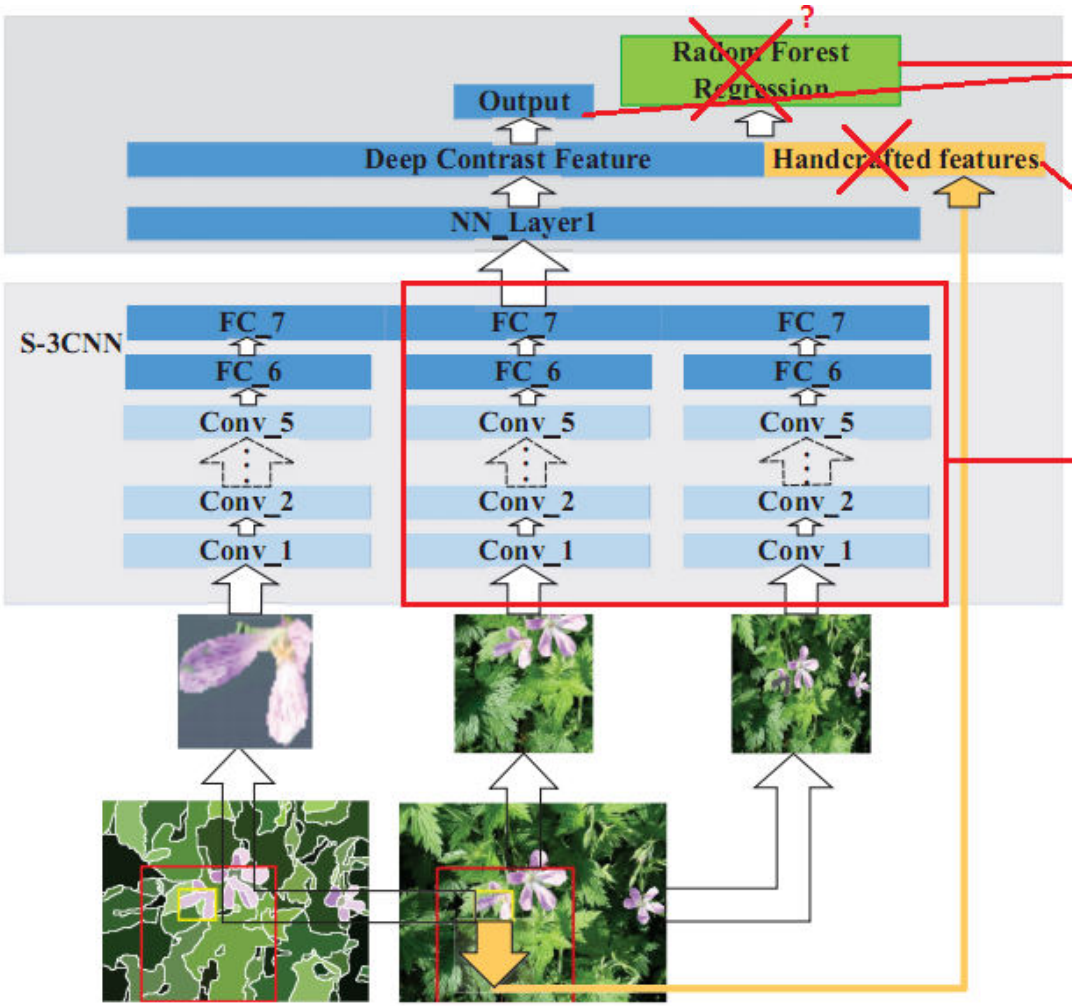
PCA для уменьшения размерности вектора Фишера

- Размерность вектора Фишера = 2 x K x D, K – количество кластеров в GMM, D – размерность вектора - локального дескриптора
- Для K = 64, D = 512 размерность вектора Фишера = 65536
- Для уменьшения размерности вектора без значительной потери качества можно воспользоваться методом PCA

Использование CNN и пулинга вектором Фишера для сегментации текстурный изображений



Модификация метода обнаружения инородных объектов



3. Если возможно, отказаться от использования дополнительных регрессоров. Проанализировать результаты, получаемые с использованием только лишь нейронной сети

2. Отказаться от использования низкоуровневых признаков, конструируеммых "вручную"

1. Вместо векторов, полученных на выходе предпоследнего полностью соединенного слоя, использовать вектор Фишера, получаемый в результате пулинга векторов с последнего сверточного слоя, рассматриваемых как локальные характеристики изображения

Использованные публикации

1. Cimpoi, M., Maji, S., Kokkinos I., Vedaldi A.: Deep filter banks for texture recognition, description, and segmentation. CoRR abs/1507.02620 (2015) URL: https://arxiv.org/abs/1507.02620
2. Guanbin Li, Yizhou Yu: Visual Saliency Based on Multiscale Deep Features. CoRR abs/1503.08663 (2015) URL: https://arxiv.org/abs/1503.08663