Свёрточные сети для семантической сегментации изображений

Михаил Фигурнов
ВМК МГУ
2014

Семантическая сегментация изображений

- Дано: RGB изображение
- Найти: метку класса для каждого пикселя

Традиционный способ решения (shallow learning)

Плюсы:

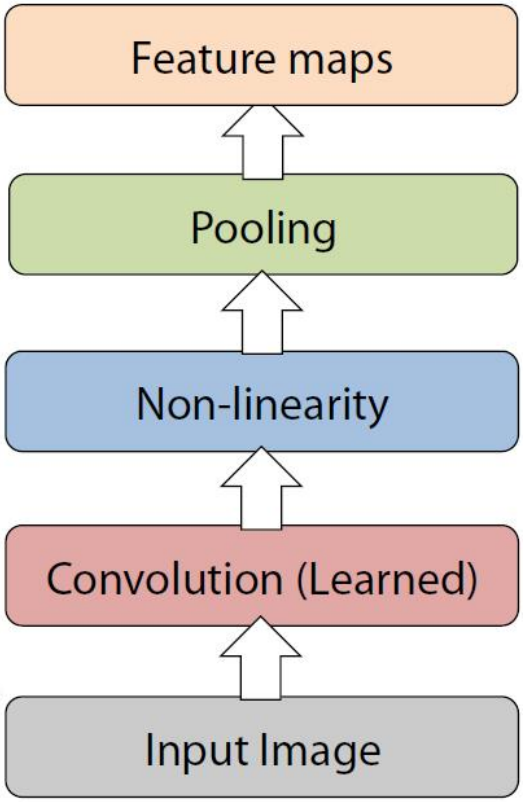
- Хорошее качество
- Понятно, как расширять

Минусы:

- Медленный вывод
- Много ручной работы
- Нельзя использовать большие данные

Convolutional Networks (LeCun et al., 89)

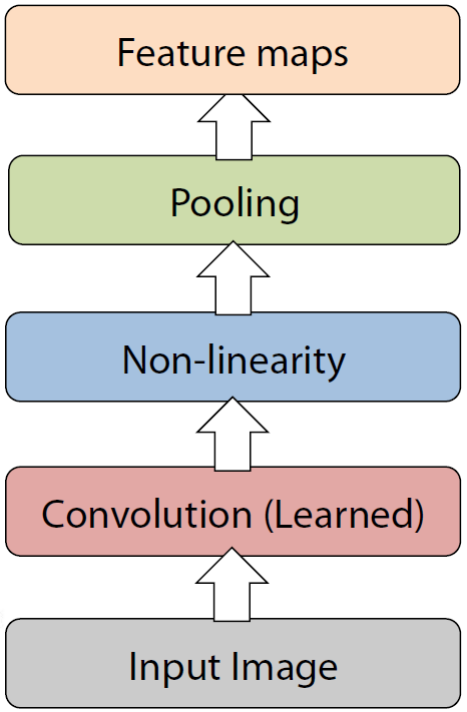
- Сеть прямого распространения
- Операции:
  - Свёртка
  - Нелинейность
  - Pooling
- Обучение с учителем
- Стохастические градиентные методы
- Градиент вычисляется методом обратного распространения ошибок (back propagation)



Структура слоя сети

- Вход: K_i-канальное изображение z
- Выход: K_{i+1}-канальное изображение y
- Свёртка (convolution)

$$y_c := \sum_{k=1}^{K_i} g_{k,c}(z_k \oplus f_{k,c})$$



- Нелинейность
  - rectified linear $y_c := \max(y_c, 0)$
  - локальная нормировка

- Pooling
  - 3D max-pooling
  - subsampling

$$y_{c,p} := \max_{\substack{d \in \text{some layers} \\ q \in \text{some pixels}}} y_{d,q}$$

Deep learning

Плюсы:

- Больше данных – выше качество
- Меньше ручной работы
- Достаточно быстрое тестирование (real-time!)

Минусы:

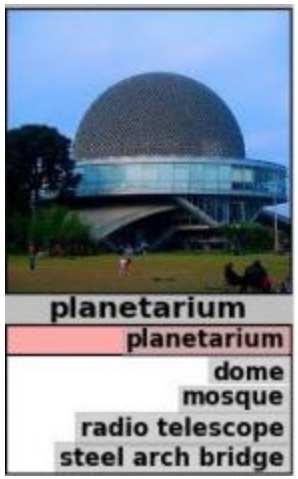
- Сложно обучить
- Не ясно, как выбирать архитектуру


DATA
is the new oil

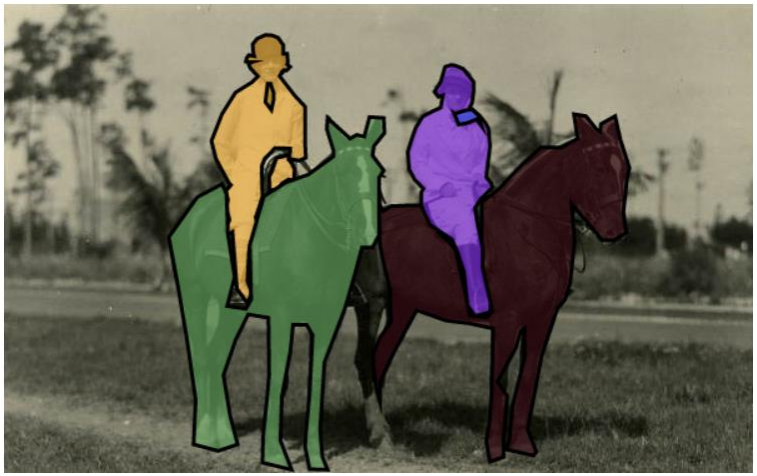
Выборки

- Для классификации –
  15 миллионов изображений

IMAGENET

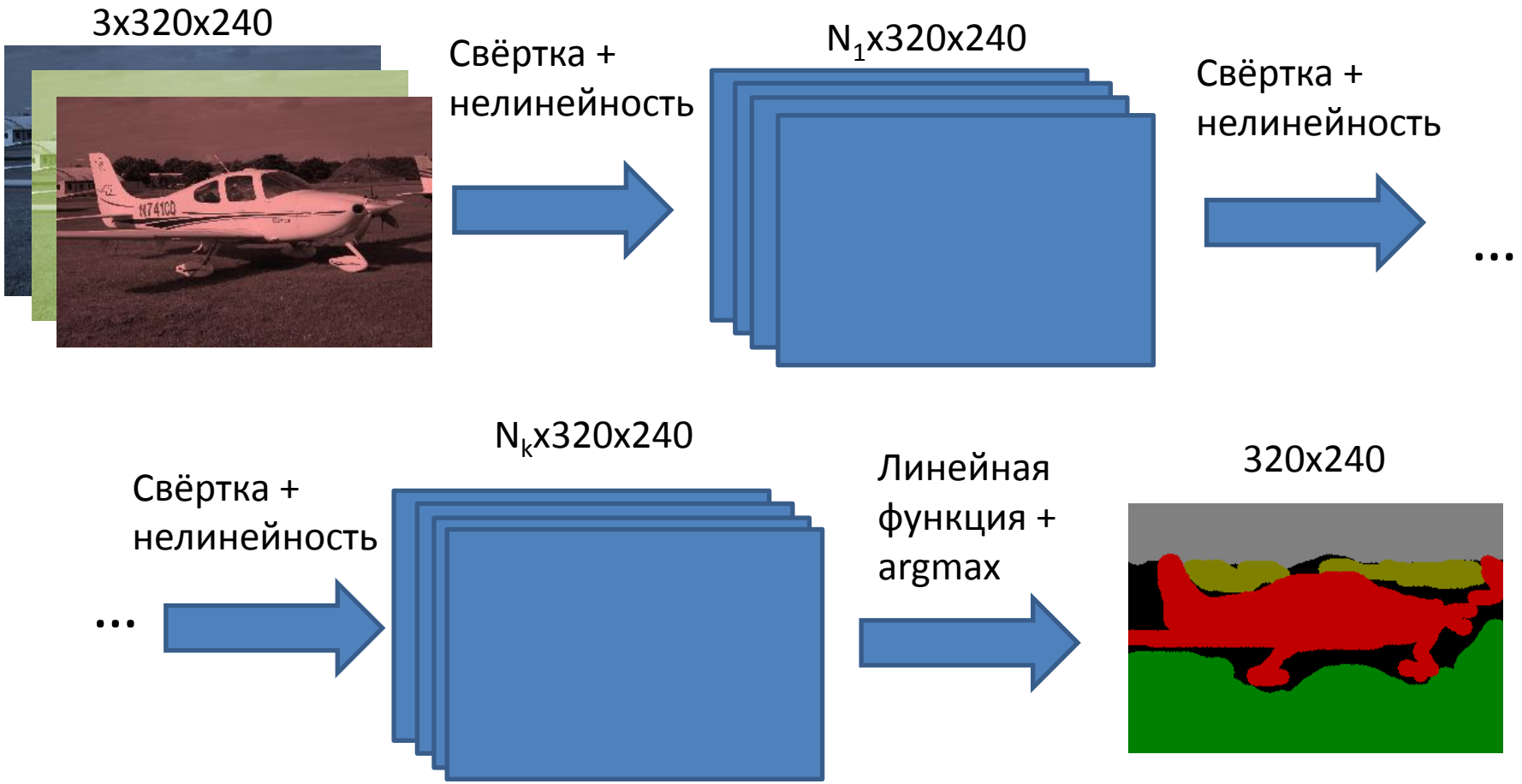


- Для сегментации
  - PASCAL VOC (2008-2012)
    10 000 изображений
  - **Microsoft COCO (2014)**
    300 000 изображений

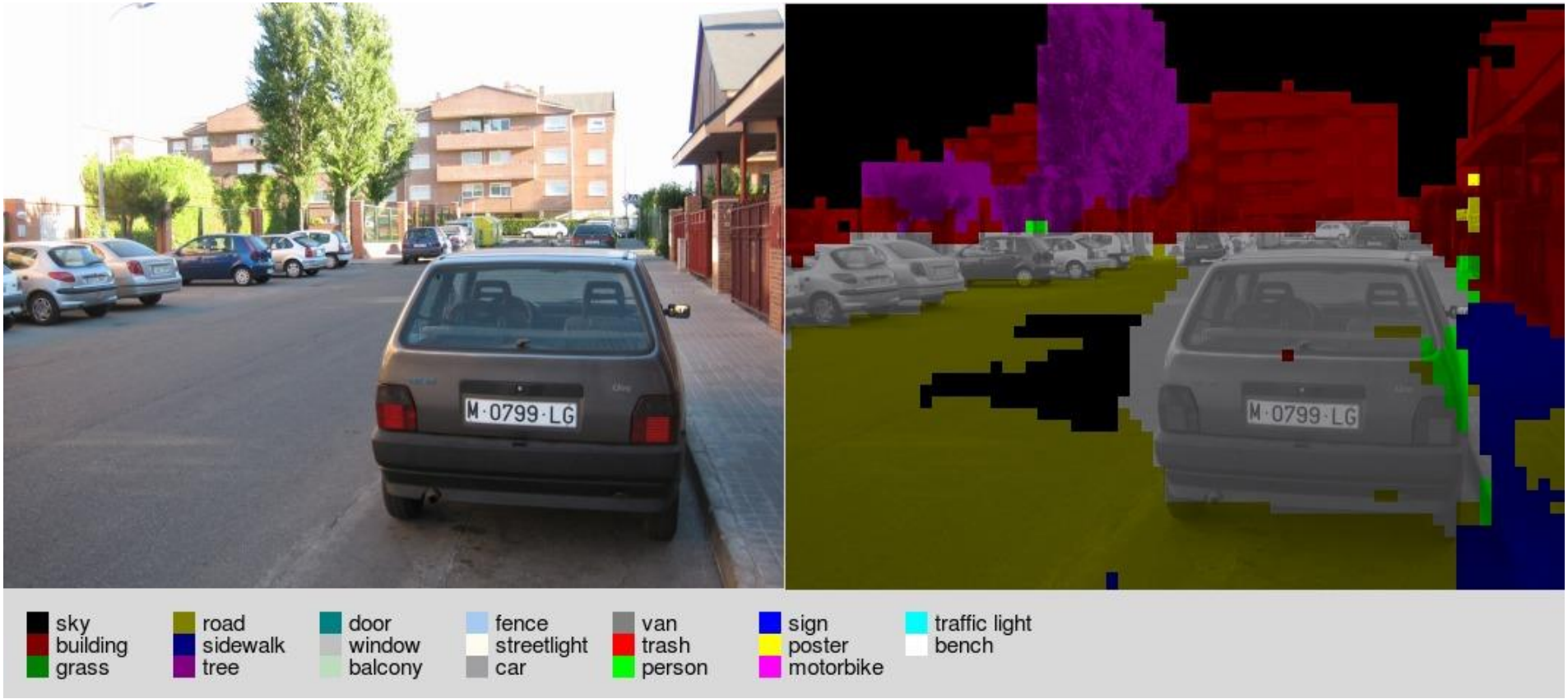


Простейший подход к сегментации

Grangier, Deep Convolutional Networks for Scene Parsing, 2009

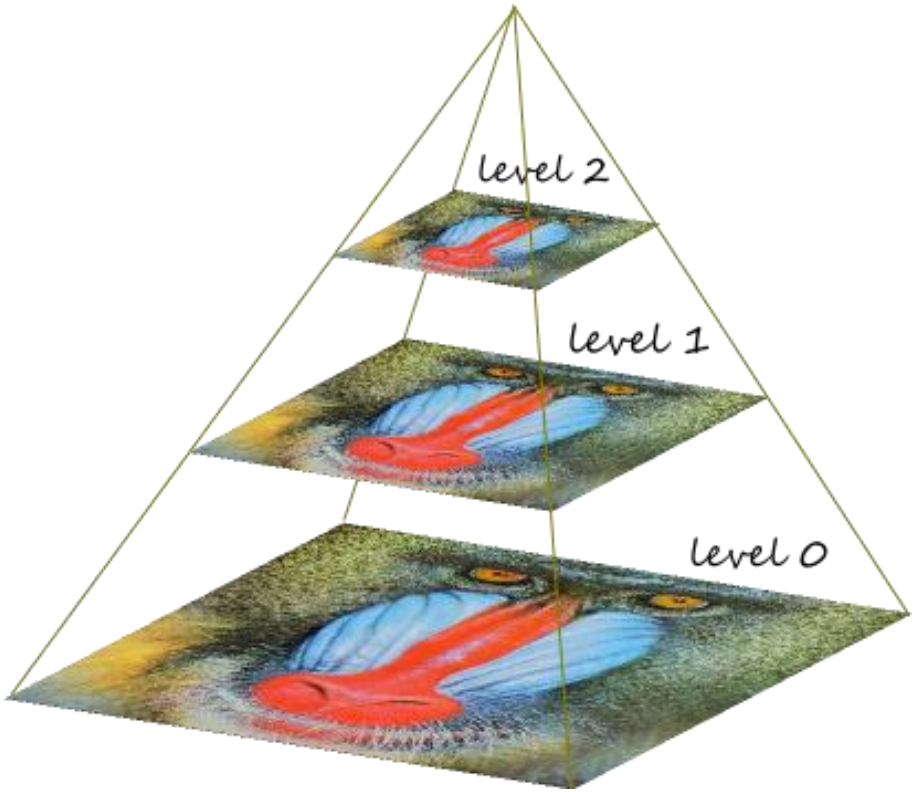


Простейший подход к сегментации - результаты



Инвариантность к масштабу

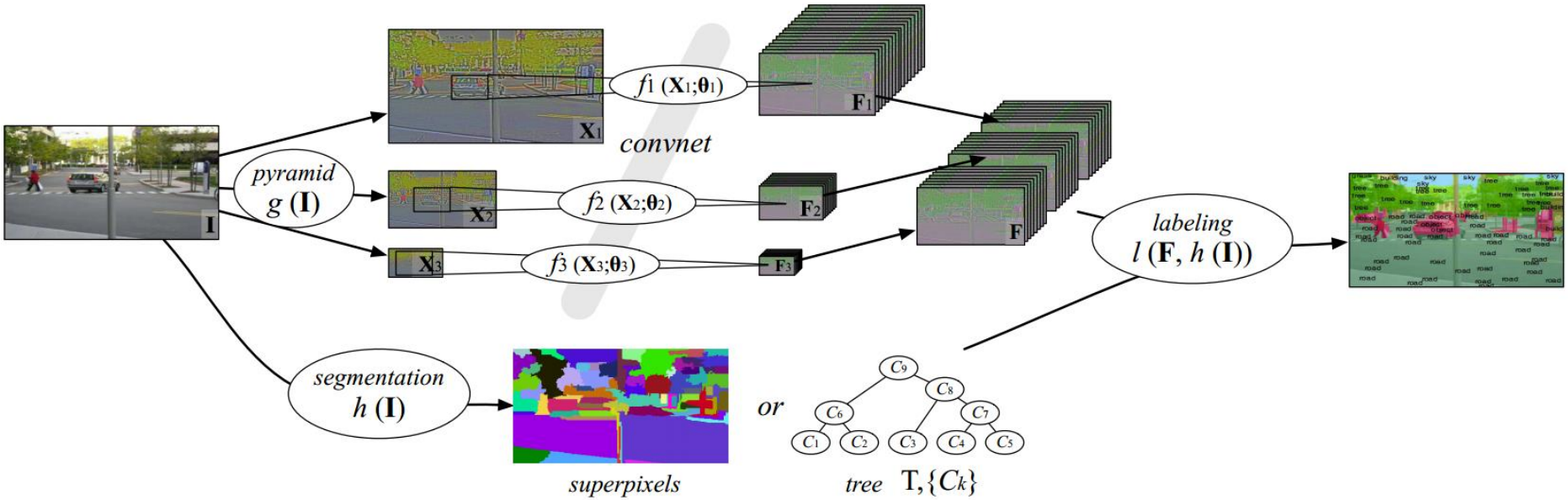
Пирамида изображений



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

Многомасштабные признаки

Farabet, Learning Hierarchical Features for Scene Labeling, 2013



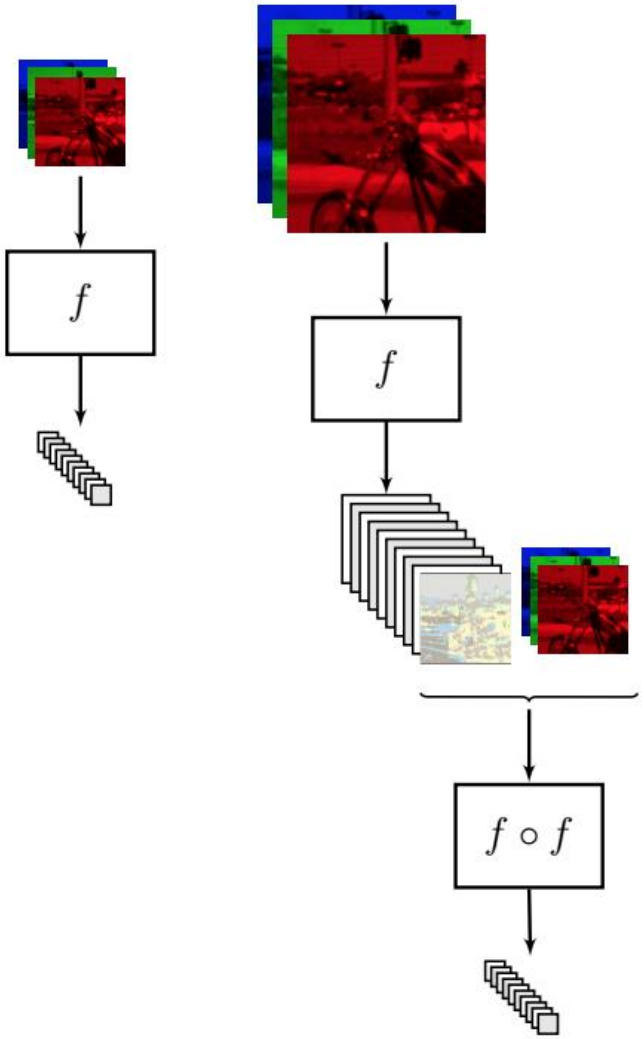
Многомасштабные признаки - результаты

Farabet, Learning Hierarchical Features for Scene Labeling, 2013

Эксперимент на Stanford Background Dataset

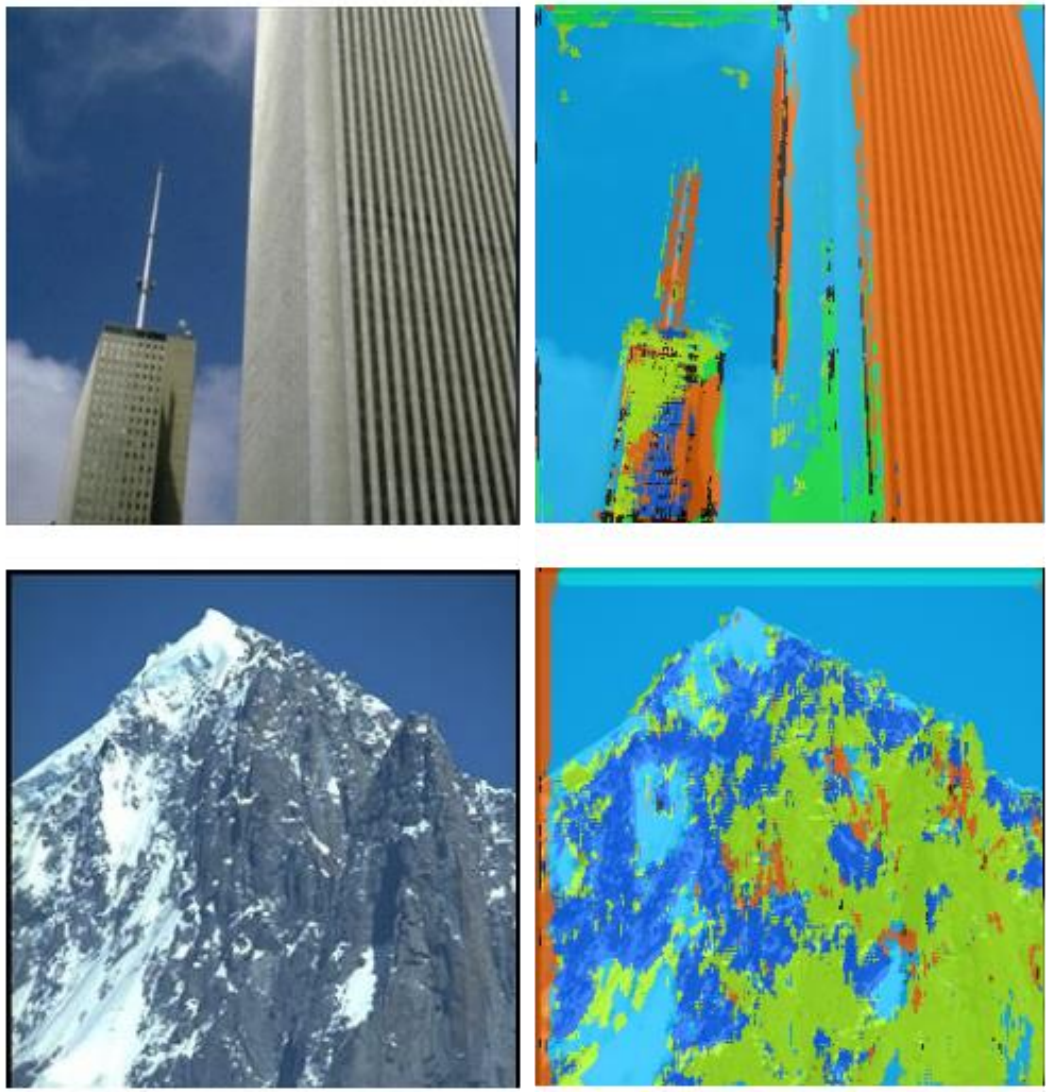
Алгоритм	Попиксельная точность	Время сегментации
State-of-the-art	81.9%	>60 секунд
Свёрточная сеть, один масштаб	66.0%	0.35 секунды
Свёрточная сеть, три масштаба	78.8%	0.6 секунды

Рекуррентные свёрточные сети



- Pinheiro, Recurrent convolutional neural networks for scene parsing, 2014
- Вход – RGB картинка, плюс вероятности за классы с прошлой итерации (на первой итерации - нули).
- Разрешение выхода меньше, чем разрешение входа!

Рекуррентные свёрточные сети - результаты



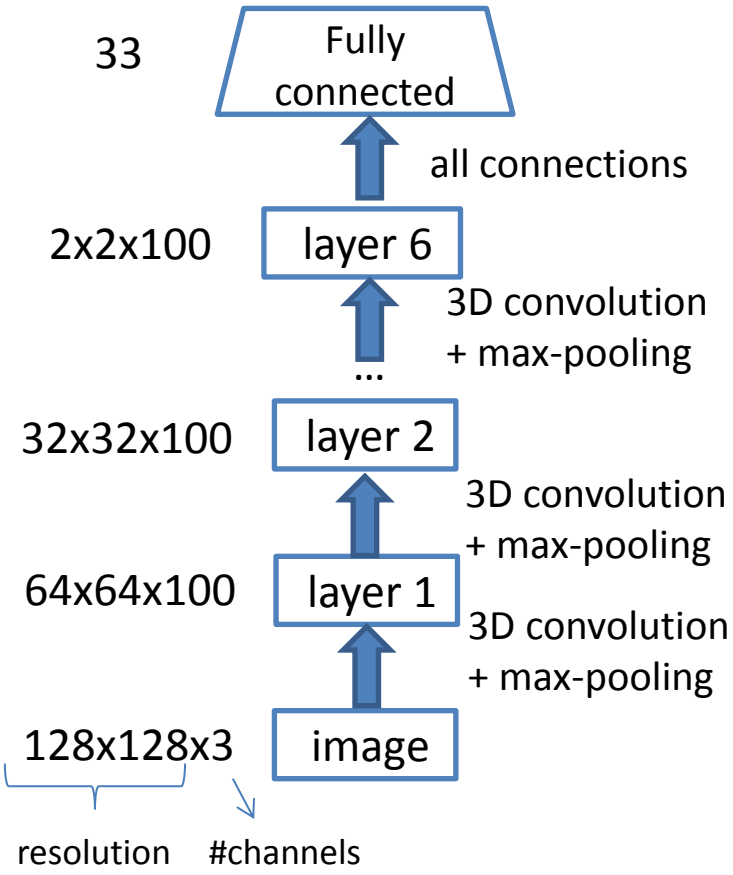
Рекуррентные свёрточные сети - результаты

Эксперимент на Stanford Background Dataset

Алгоритм	Попиксельная точность	Время сегментации
Свёрточная сеть, три масштаба	78.8%	0.6 секунды
Рекуррентная сеть, разрешение выхода 1/8	78.4%	0.2 секунды
Рекуррентная сеть, разрешение выхода 1/4	79.3%	0.7 секунды

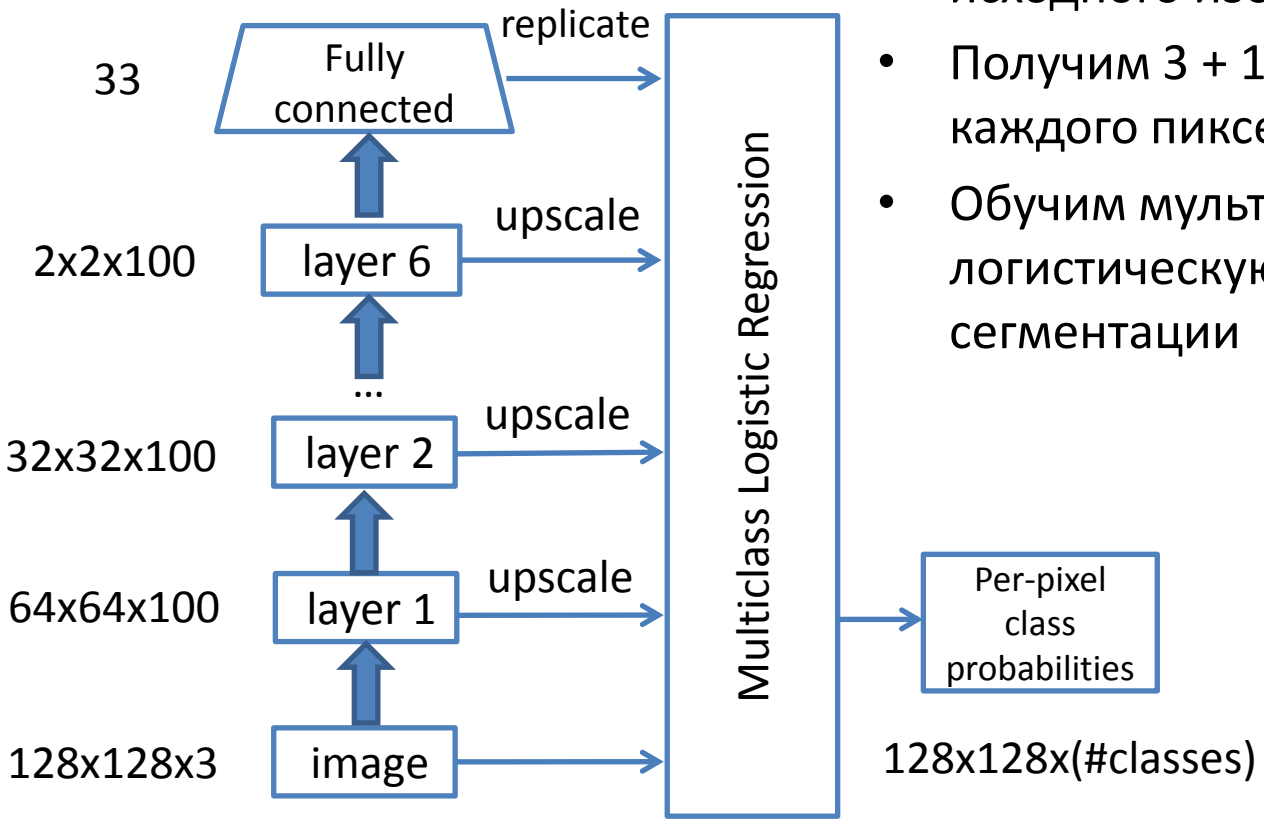
Улучшаем обратную связь

Gatta, Unrolling loopy top-down semantic feedback in convolutional neural networks, 2014



- Выучим при помощи unsupervised метода «пирамидальную» свёрточную сеть
- Разные уровни признаков: от пикселей до всего изображения
- Верхний слой соответствует всему изображению

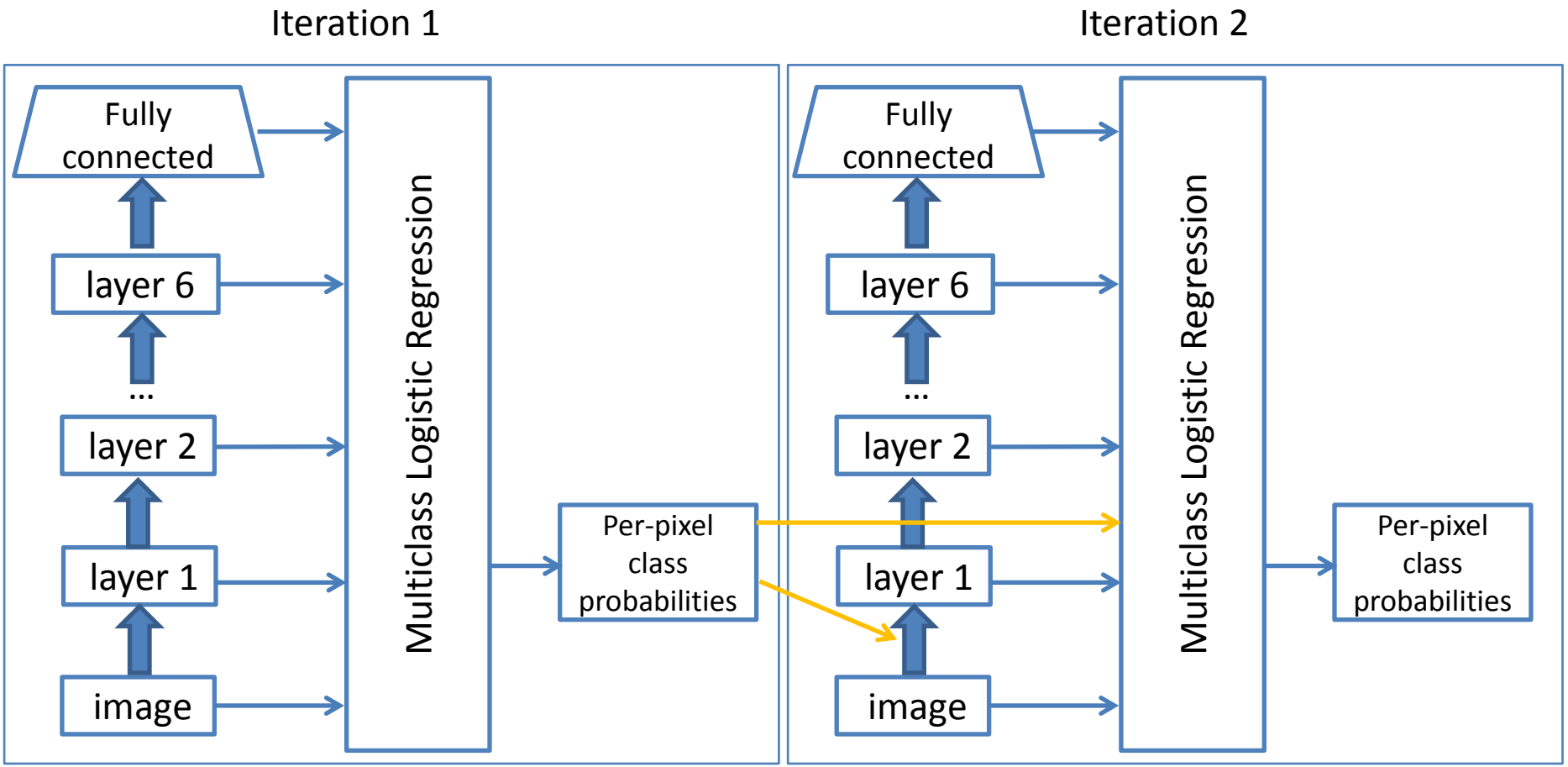
Улучшаем обратную связь



- Масштабируем все слои до размера исходного изображения
- Получим 3 + 100*6 + 33 признаков для каждого пикселя.
- Обучим мульти-классовую логистическую регрессию для сегментации

Улучшаем обратную связь

- Используем карты вероятностей от предыдущей сети как вход для следующей.
- Параметры на разных итерациях различные!

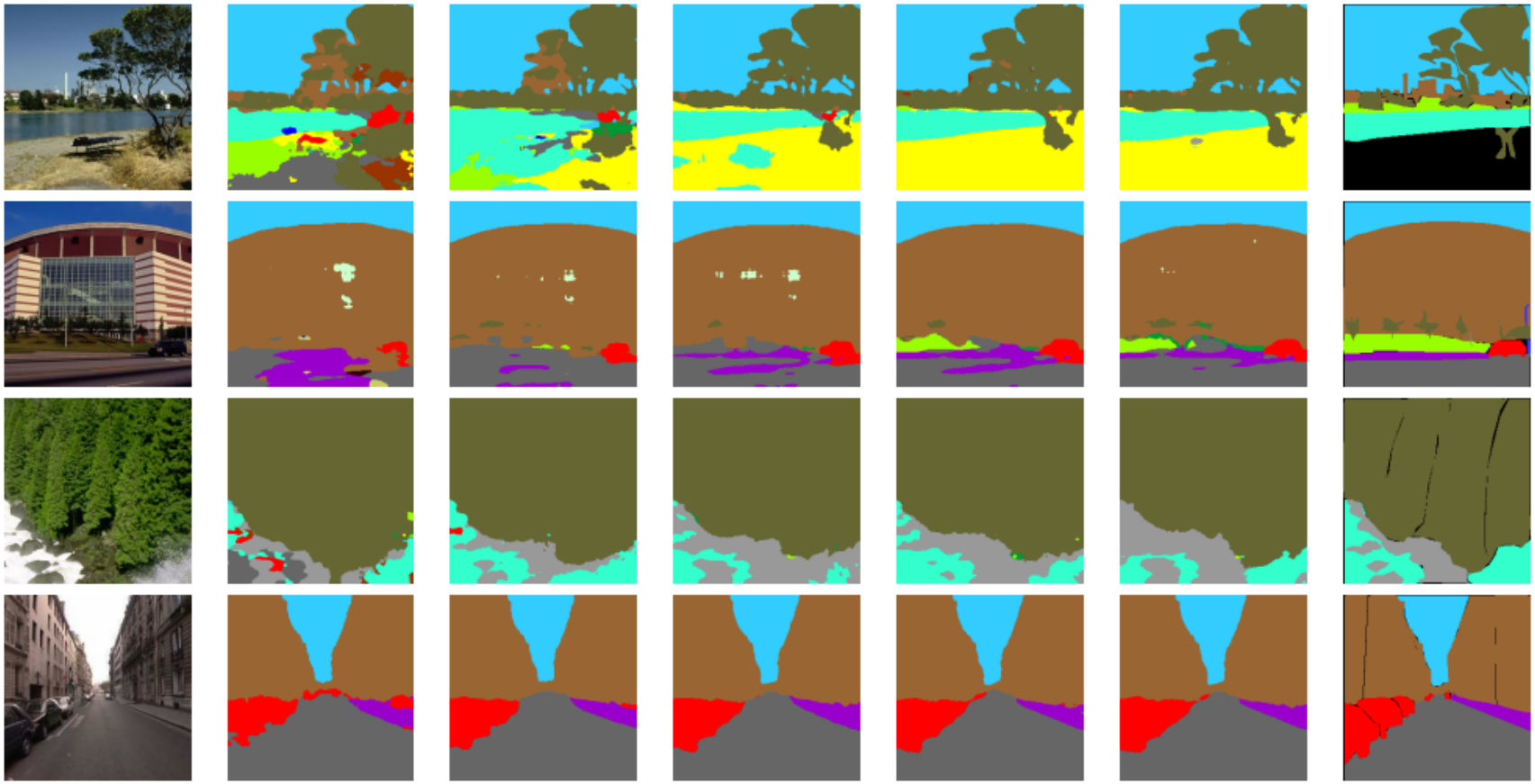


Улучшаем обратную связь - результаты

- Быстрое обучение: 4 часа против 1 недели для рекуррентных сетей
  (за счёт unsupervised обучения)
- Почти state-of-the-art качество на SiftFlow Dataset

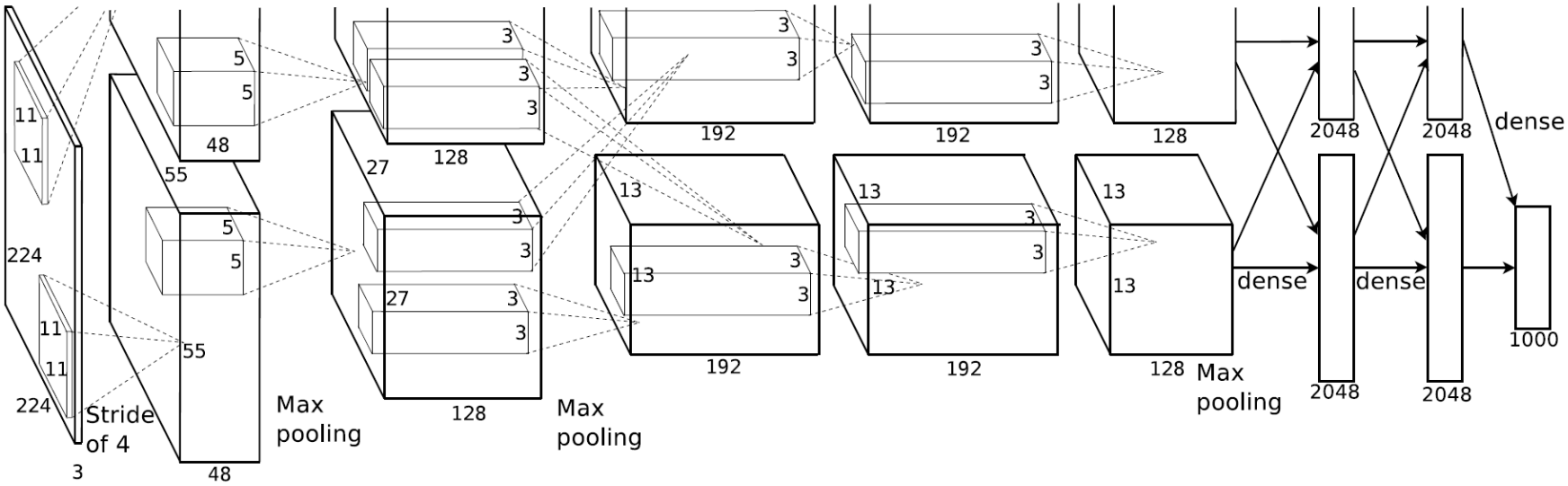
Алгоритм	Попиксельная точность
Рекуррентная свёрточная сеть	77.7%
1 итерация	74.7%
5 итерация	78.7%

Улучшаем обратную связь - результаты



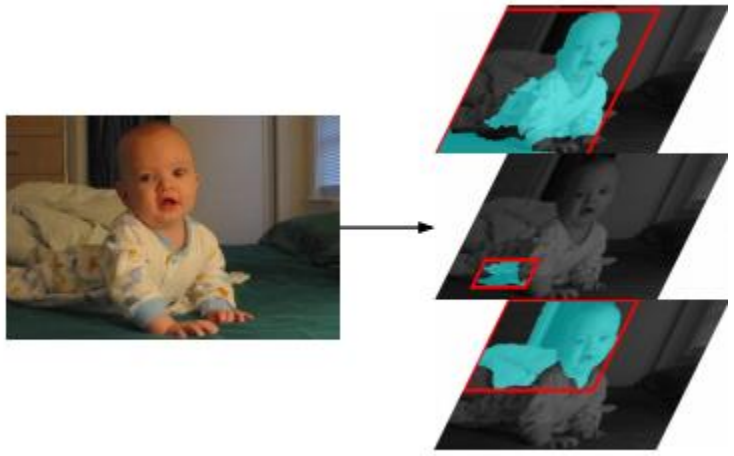
Архитектура для ImageNet (2012)

- 1,000,000 объектов
- 224 x 224 пикселей
- 1,000 классов



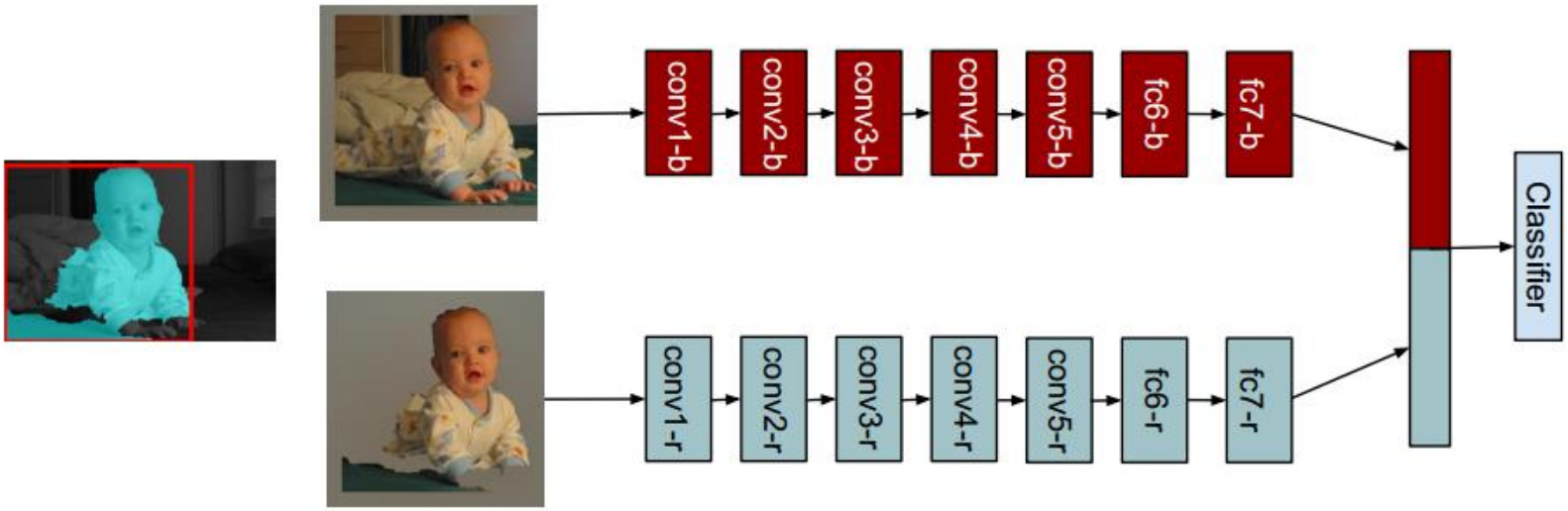
Классификация регионов

- Hariharan, Simultaneous Detection and Segmentation, 2014
- Используем метод MCG для поиска возможных регионов на изображениях

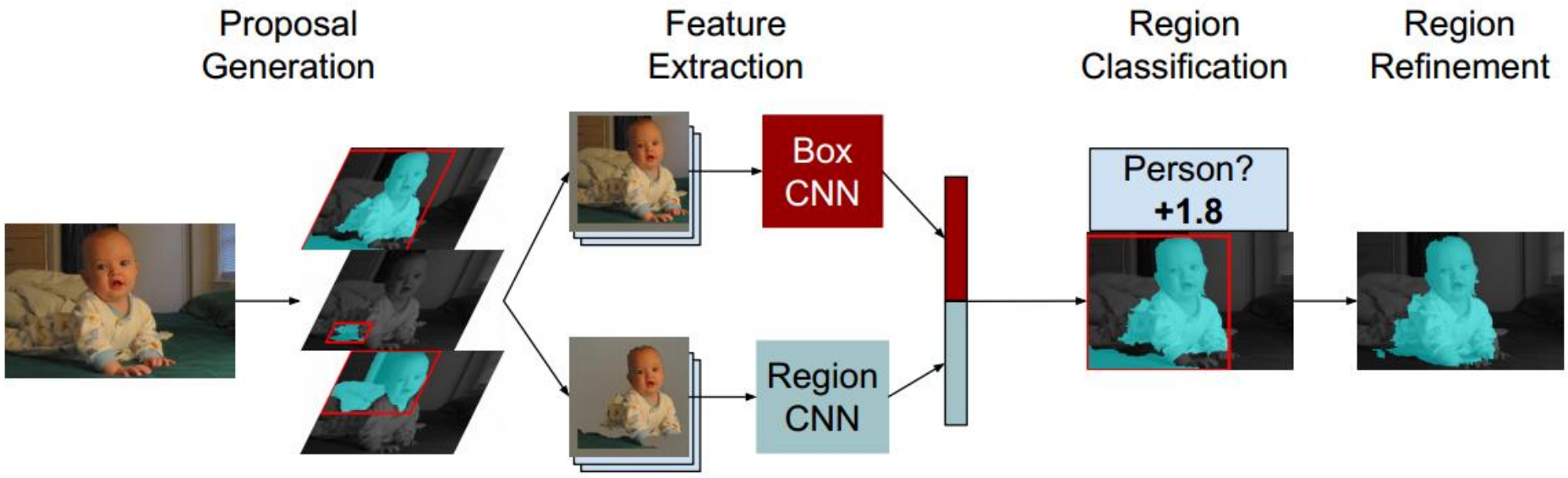


Классификация регионов – обучение

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



Классификация регионов - тестирование



Классификация регионов - результаты

- Находит отдельные объекты
- Высокое качество
- Не полностью deep learning (можно лучше?)

Слабое обучение

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

Кошки

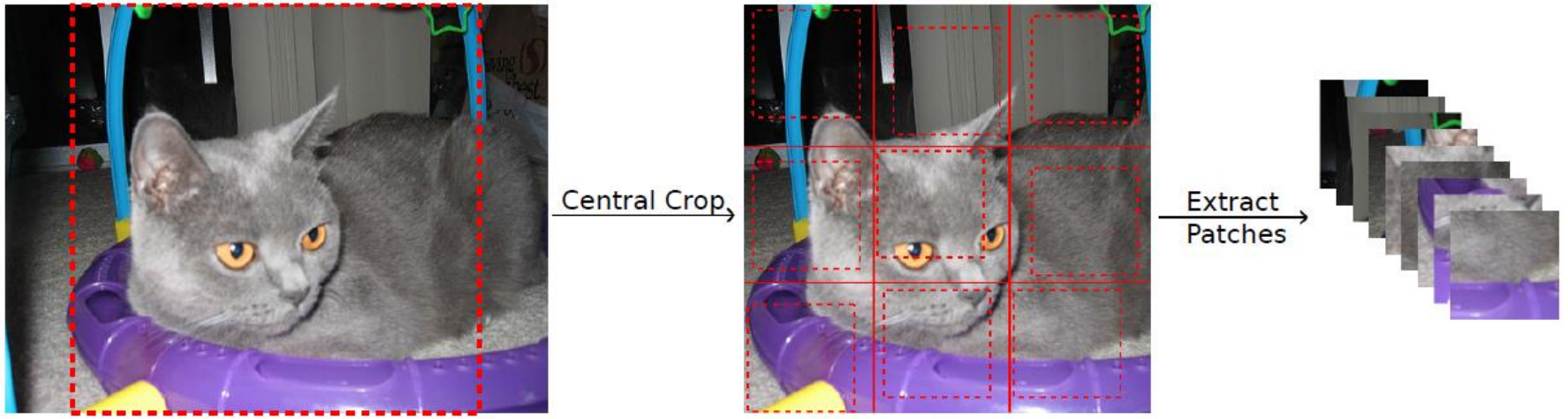
37345 изображений

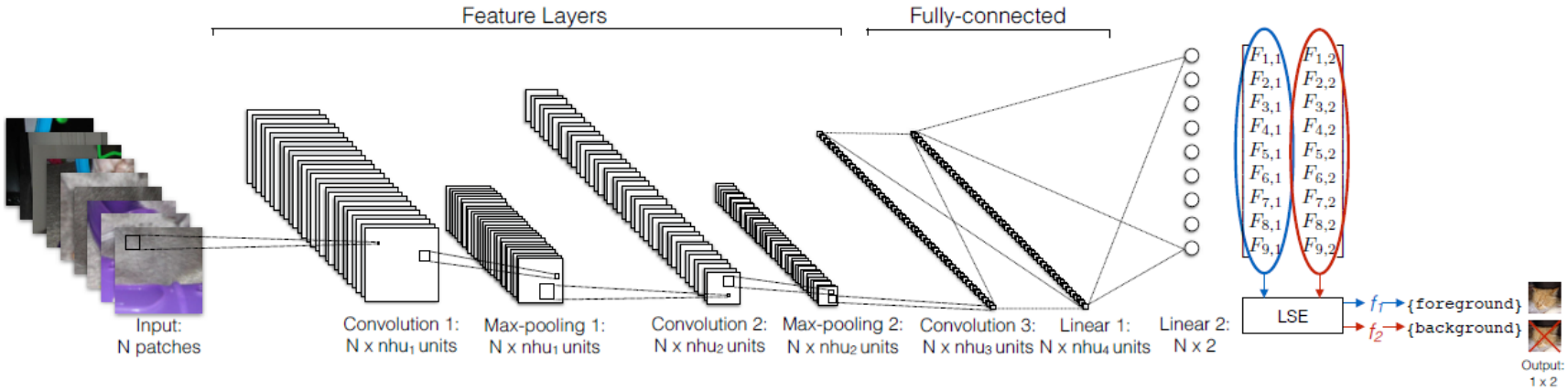
Собаки

187775 изображений

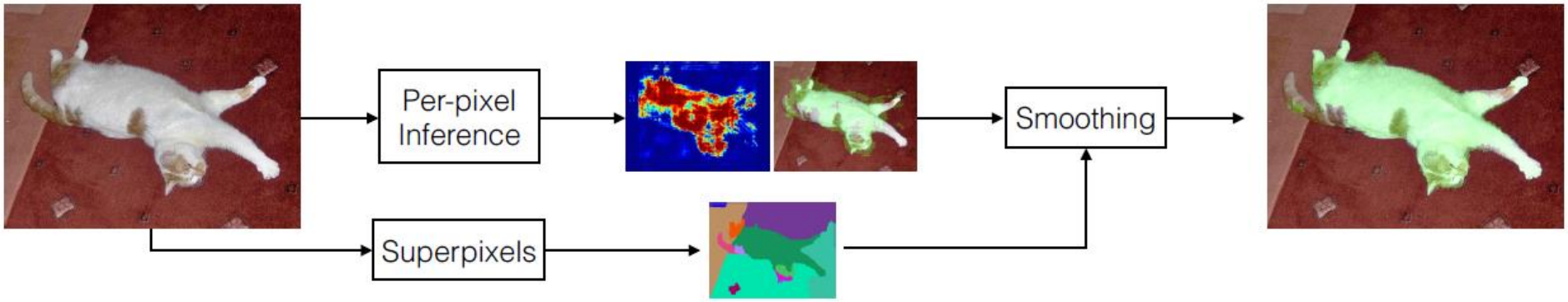
IMAGENET

Слабое обучение – патчи

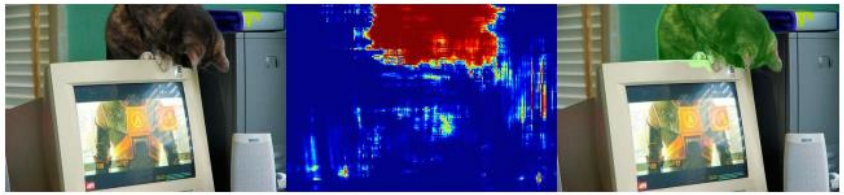


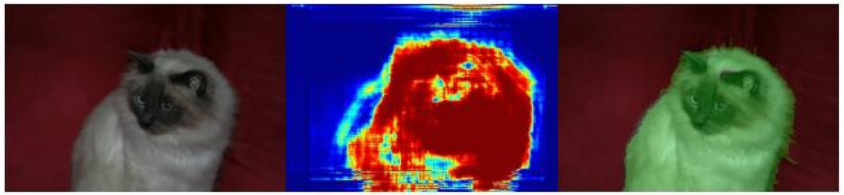


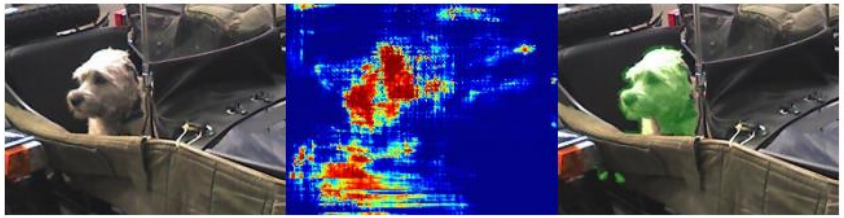
Слабое обучение - тестирование

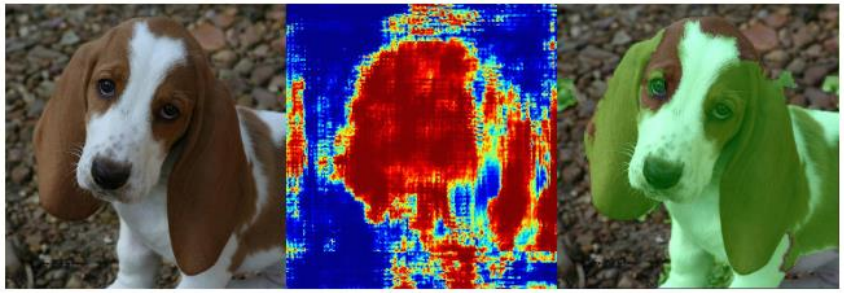


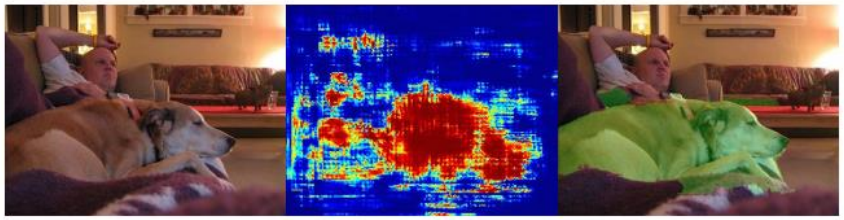
Слабое обучение - результаты











Выводы

- Свёрточные сети можно использовать для семантической сегментации множеством способов
- Можно добиться как высокой скорости работы, так и высокого качества
- Для deep learning нужно много данных
- С появлением больших выборок можно будет обучить более сложные архитектуры