



# **О задаче поиска объектов на изображении**

Евгений Борисов

# ML обработка изображений

## цели обработки изображений

**Полиграфия, дизайн** — улучшение качества, ретушь, изменение размера и формы, композиция.

**Спецэффекты в кино** – композиция, монтаж фонов, захват движения.

**Интернет** — поиск, аннотация, поиск дубликатов, распознавание объектов.

**Промышленные системы** — диагностика, контроль качества.

**Роботы и видеонаблюдение** — поиск и локализация объектов, отслеживание, распознавание объектов, распознавание жестов и событий.

# ML обработка изображений

## задачи обработки изображений

сравнение изображений

коррекция изображения

стилизация изображений

генератор изображений

классификатор изображений

сегментация изображения

локализация объектов

сопровождение объектов

реконструкция сцены

распознавание событий

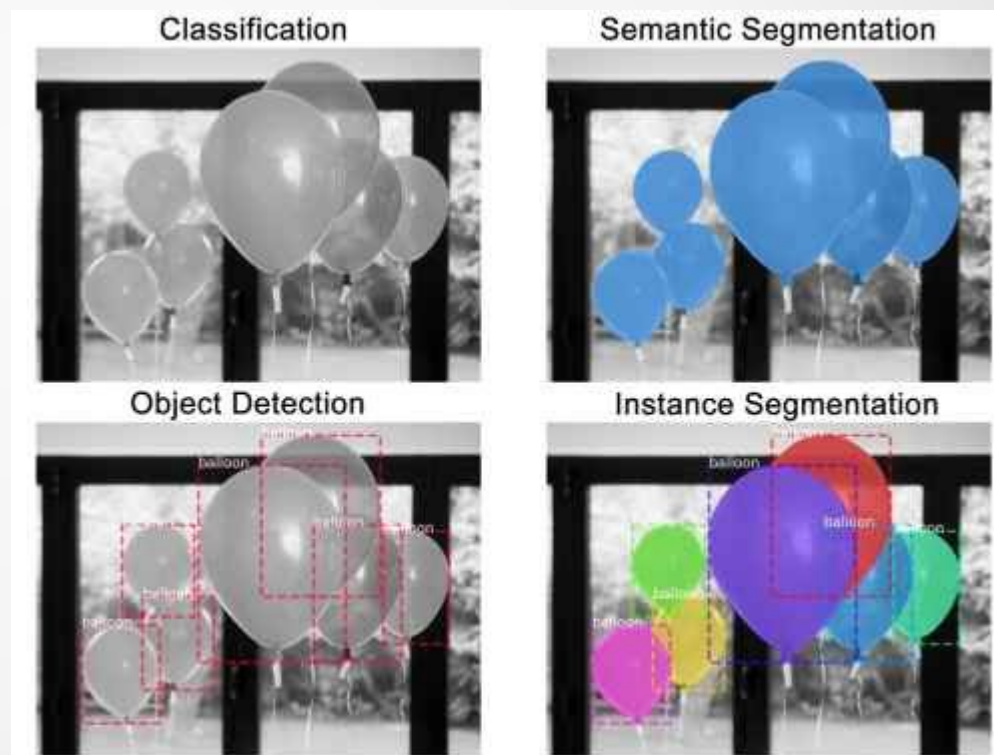
# Задача поиска объектов на изображении

**Classification** — классификация изображения по типу объекта, которое оно содержит;

**Semantic segmentation** — определение всех пикселей объектов определённого класса или фона на изображении. Если несколько объектов одного класса перекрываются, их пиксели никак не отделяются друг от друга;

**Object detection** — обнаружение всех объектов указанных классов и определение охватывающей рамки для каждого из них;

**Instance segmentation** — определение пикселей, принадлежащих каждому объекту каждого класса по отдельности;



# Поиск объектов на изображении

## Модель объекта

- необходимо понимать что хотим найти

Цветовые фильтры

Выделение и анализ контуров

Сопоставление с шаблоном

Работа с особыми точками

Методы машинного обучения

## Модель фона

- камера неподвижна

Усреднённый фон

Модель фона по Гауссу

Смесь гауссиан как модель фона

# Поиск объектов на изображении

## **Модель фона**

- камера неподвижна

Усреднённый фон

Модель фона по Гауссу

Смесь гауссиан как модель фона

# Поиск объектов на изображении

## Модель фона / Усреднённый фон

Соберём историю из  $n$  кадров,  
вычислим среднее значение  
и будем использовать этот результат как фон

$$B = \frac{1}{n} \sum_{i=1}^n C_i$$

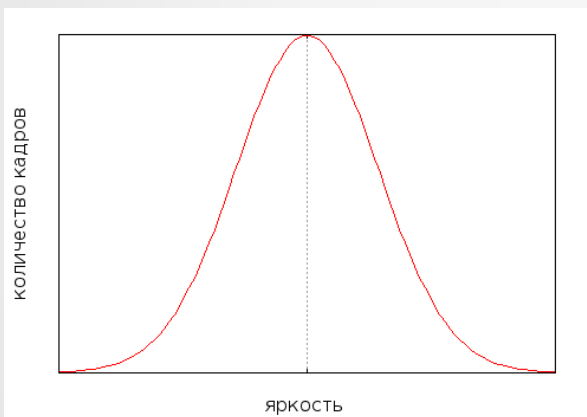


работает удовлетворительно только в условиях  
стабильного освещения и отсутствия шума

# Поиск объектов на изображении

## Модель фона / Модель фона по Гауссу

состояние точек фона находится в окрестности определённого значения  
фон задаётся нормальными распределениями яркостей точек



$$\mu = \frac{1}{n} \sum_{i=1}^n x_i ; \quad \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2$$

$$p(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$





# Поиск объектов на изображении

## Модель фона / Смесь гауссиан

фон задаётся смесью нормальных распределений яркостей точек  
позволяет формировать несколько кластеров для значений яркости

$$p(x) = \sum_{j=1}^K w_j \varphi_j(x)$$

$$\varphi(x; \mu, \Sigma) = \frac{\exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)}{\sqrt{(2\pi)^n \det \Sigma}}$$



устойчивей к шуму и изменениям освещения чем простые модели

# Поиск объектов на изображении

## **Модель объекта**

- необходимо понимать что хотим найти

Цветовые фильтры

Выделение и анализ контуров

Сопоставление с шаблоном

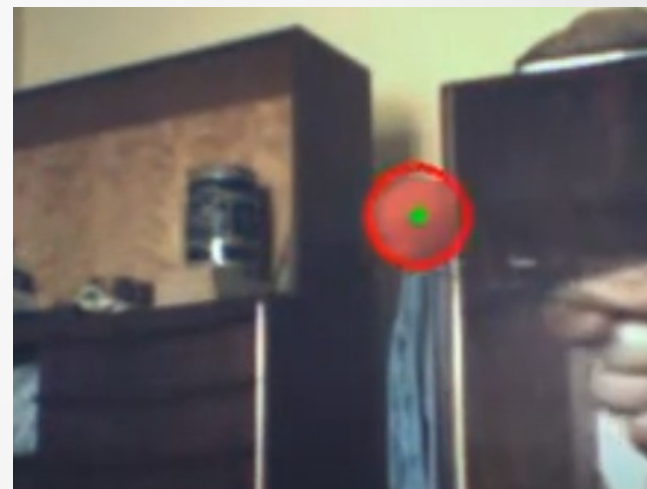
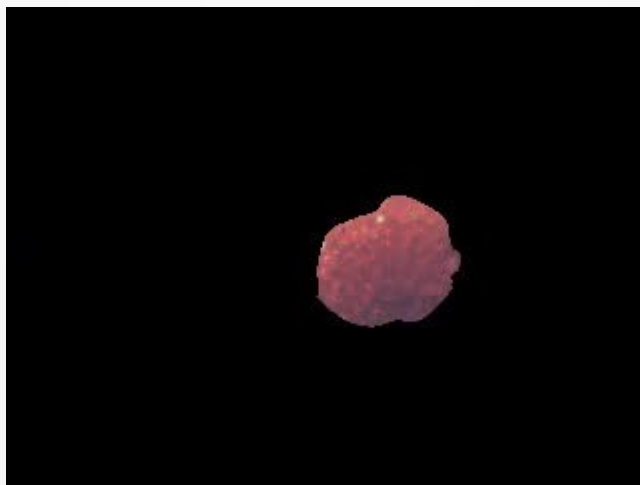
Работа с особыми точками

Методы машинного обучения

# Поиск объектов на изображении

## Модель объекта / Цветовые фильтры

объект существенно отличаться от фона по цвету,  
Имеет однородную раскраску  
освещение равномерно и не изменяется

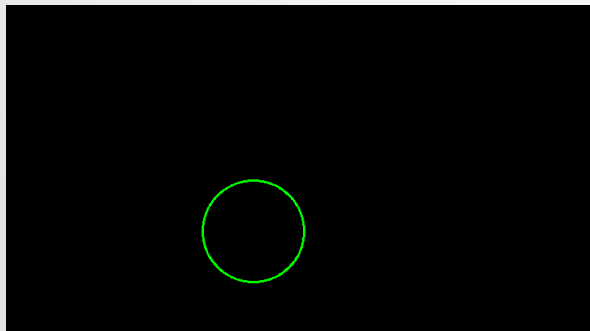
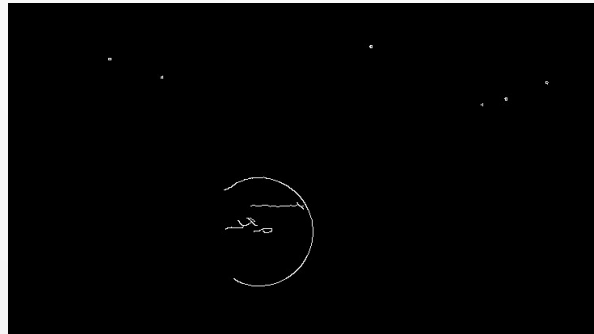


# Поиск объектов на изображении

## Модель объекта / Выделение и анализ контуров

выделяем границы на изображении (метод Canny)

проверить выделенные линии-границы на соответствие геометрическим контурам объекта (метод Хафа / Hough Transform)



## Поиск объектов на изображении

**Модель объекта** / Сопоставление с шаблоном  
ищем на большом изображении области  
совпадающие с изображением объекта



Поворот и масштабирование могут сильно портить результат



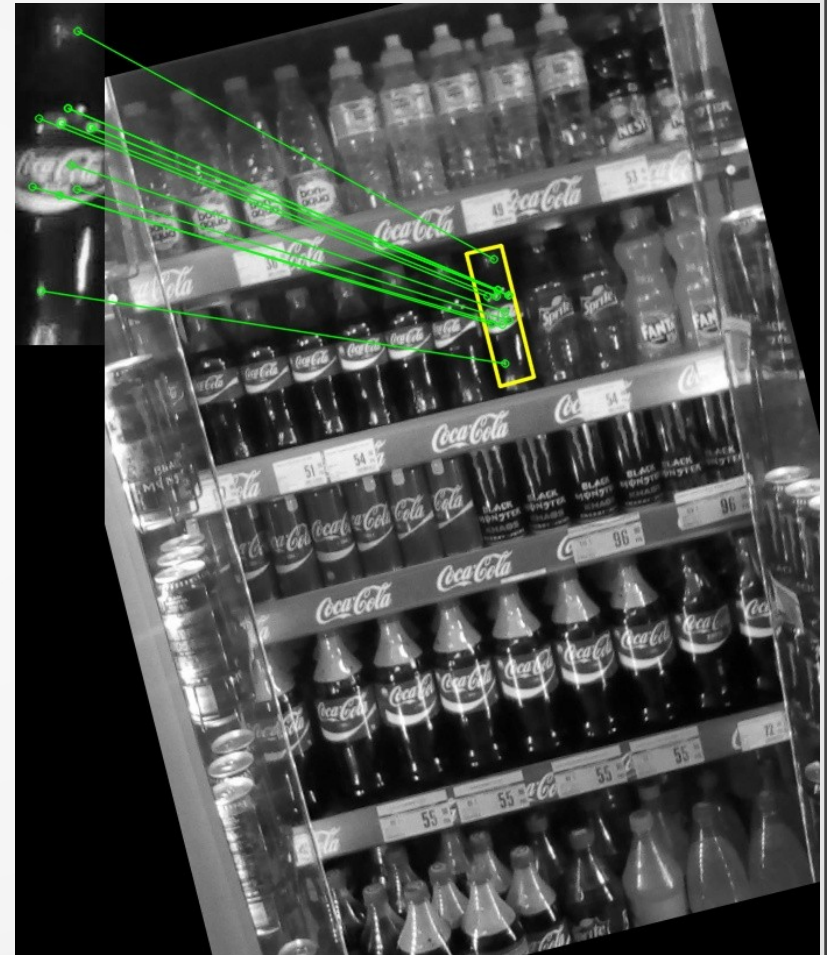
# Поиск объектов на изображении

## Модель объекта / Работа с особыми точками

особая точка - небольшая область, которая существенным образом выделяется на изображении (углы).

по окрестности особой точки вычисляют дескриптор (SIFT, SURF, ORB)

1. На картинке с объектом ищем особые точки/дескрипторы.
2. На анализируемом изображении тоже ищем особые точки/дескрипторы.
3. Сравниваем два набора дескрипторов.



# Поиск объектов на изображении

## **Модель объекта**

- необходимо понимать что хотим найти

Цветовые фильтры

Выделение и анализ контуров

Сопоставление с шаблоном

Работа с особыми точками

**Методы машинного обучения**

# ML обработка изображений

## **Модель объекта средствами ML**

методы извлечения признаков из картинки

Гистограммы направленных градиентов (HOG)

Признаки Хаара

«Визуальные» слова (BoW)

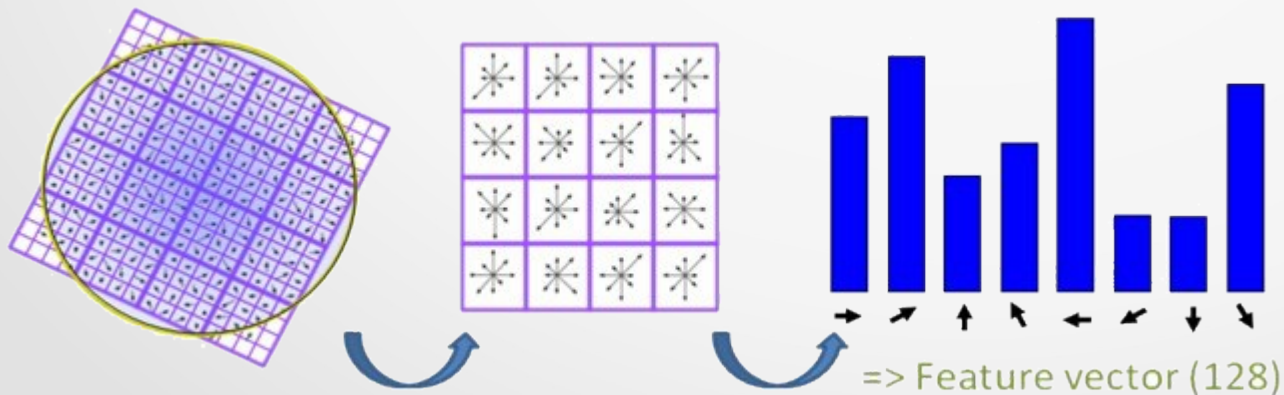
Свёрточные нейросети



# ML обработка изображений

## HOG - гистограммы направленных градиентов

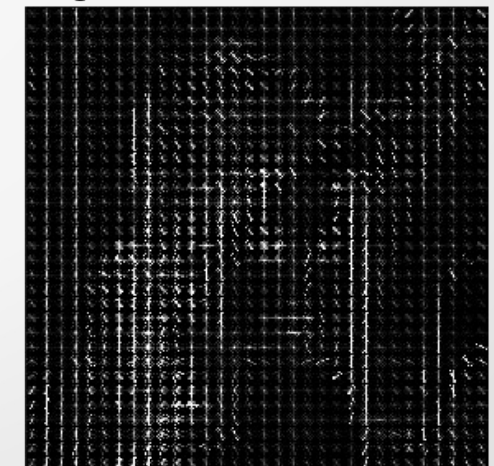
Картинка разделяется на части (ячейки),  
для каждой ячейки строим гистограмму направлений  
градиента яркости, далее гистограммы ячеек  
нормируются по контрасту и объединяются



Input image



Histogram of Oriented Gradients

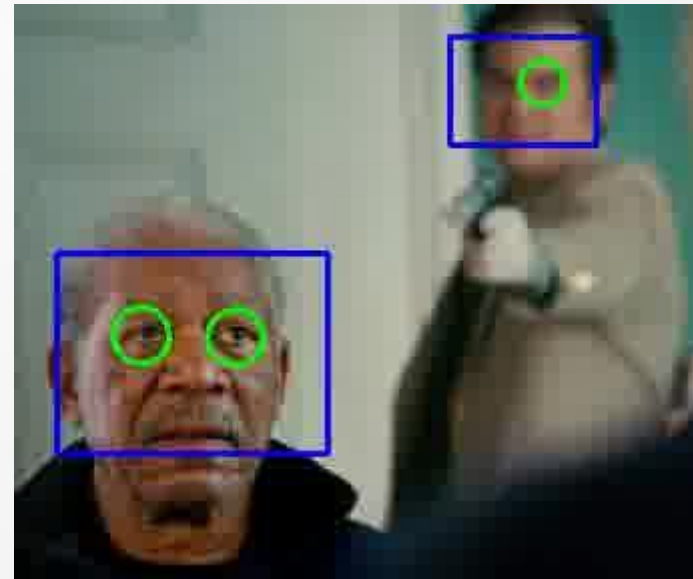


# Поиск объектов на изображении

## Извлечение признаков из картинки

### Признаки Хаара (Haar-like features)

Выбираем прямоугольную область на изображении, разбиваем её на несколько смежных прямоугольных частей, в каждой части суммируем яркость точек, вычисляем разность между этими суммами.



результат работы  
детектора лиц Виолы-Джонса

# ML обработка изображений

## Извлечение признаков из картинки

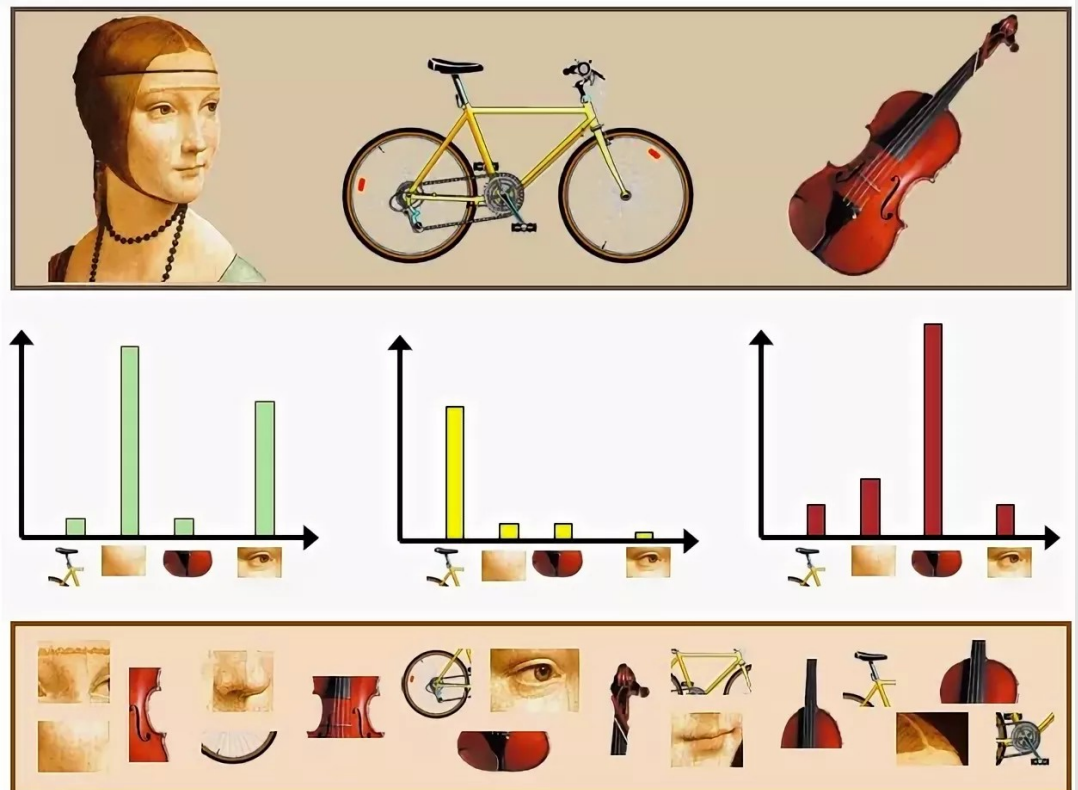
### Мешок слов (BoW)

похоже на частотный анализ текстов,

для каждого изображения, входящего в учебный набор, определяем особые точки/дескрипторы и объединяем похожие дескрипторы в группы (кластеризация)

кластер дескрипторов - «визуальное» слово

вектор BoW-признаков - количество найденных «визуальных» слов

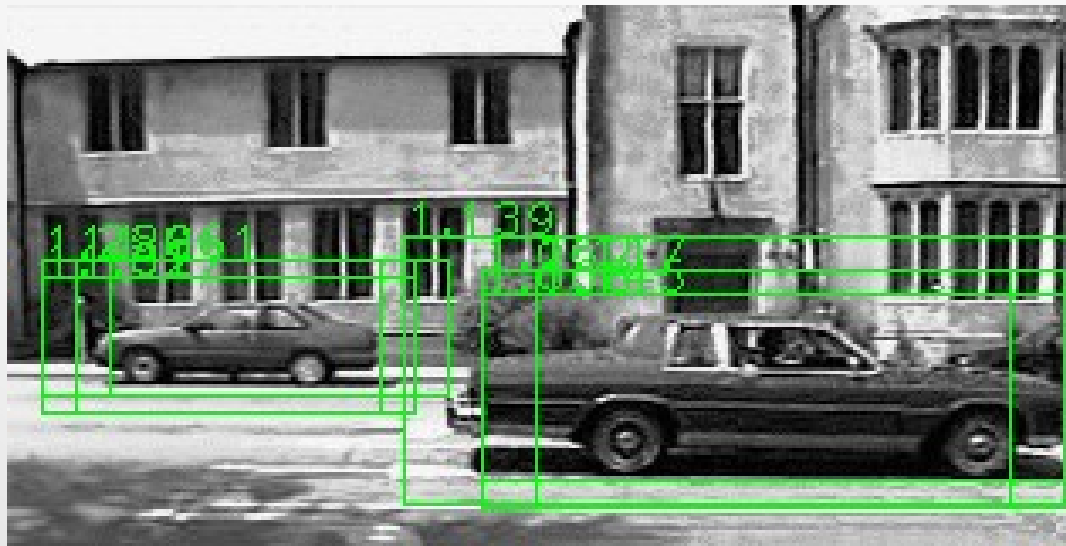


## Поиск объектов на изображении

# Методы машинного обучения и локализация объектов

## метод скользящего окна

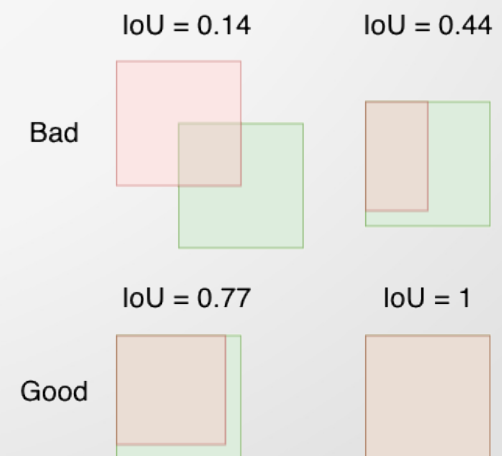
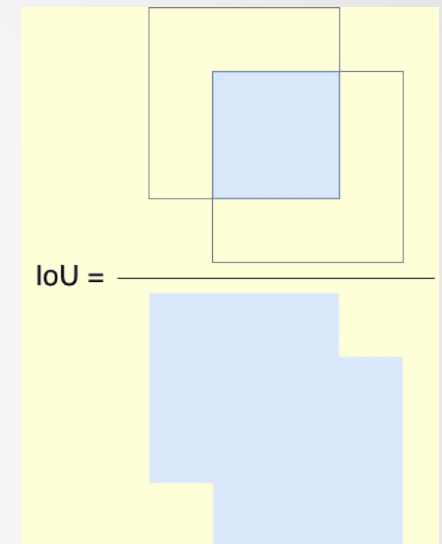
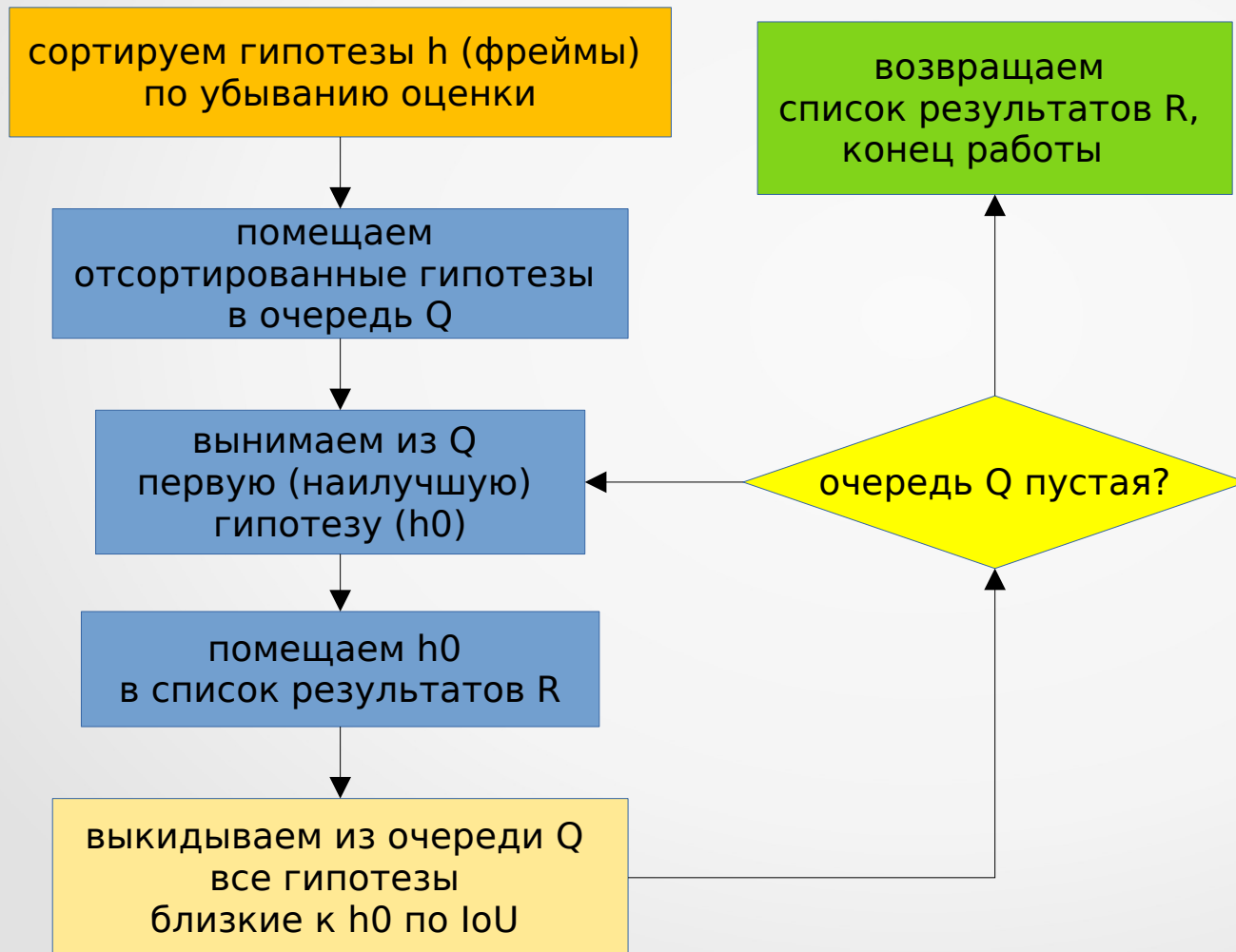
1. задать размер окна
2. пройти окном изображение
3. на каждом шаге выполняем классификацию содержимого окна
4. изменить размер окна и повторить процедуру с п.2
5. обрабатываем результаты



# Поиск объектов на изображении

## Non-maximum suppression (NMS)

оцениваем степень наложения фреймов друг на друга  
и выкидываем лишнее



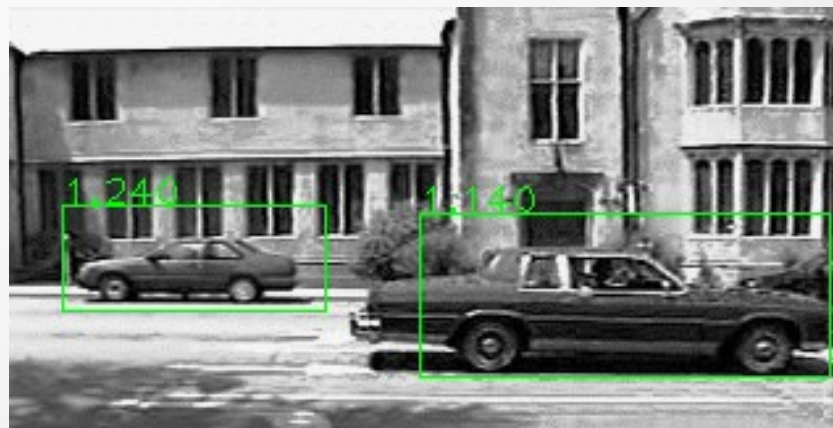


# Поиск объектов на изображении

## Локализация объектов : обработка результата

Non-maximum suppression (NMS)

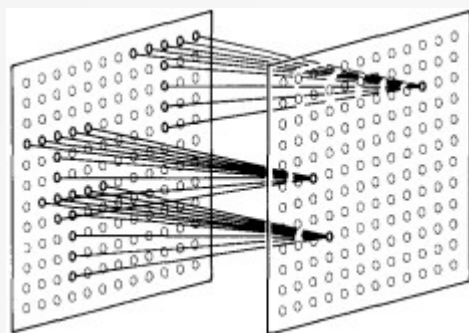
оцениваем степень наложения фреймов друг на друга и выкидываем лишнее



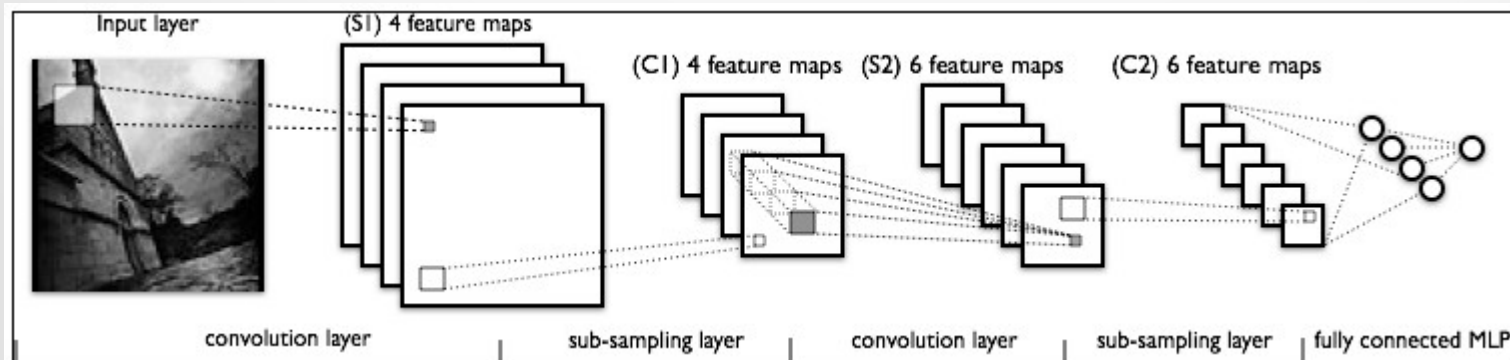
# Свёрточные нейросети

## Свёрточные сети

Fukushima, Neocognitron (1980). "A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". Biological Cybernetics. 36 (4): 193–202. doi:10.1007/bf00344251.



Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, Winter 1989.



# Поиск объектов на изображении

## Модель фона с помощью нейросети

семантическая сегментация



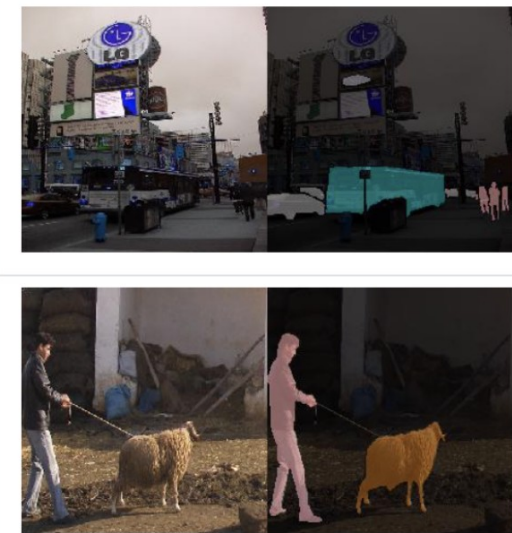
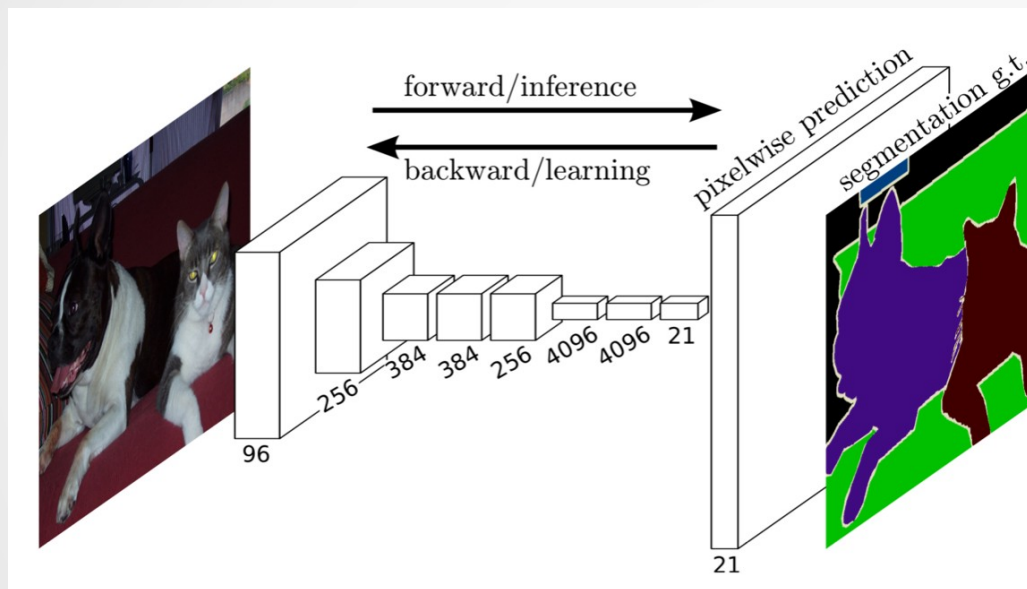


# Поиск объектов на изображении

## Semantic segmentation

FCN: Fully Convolutional Networks

<https://arxiv.org/pdf/1411.4038.pdf>



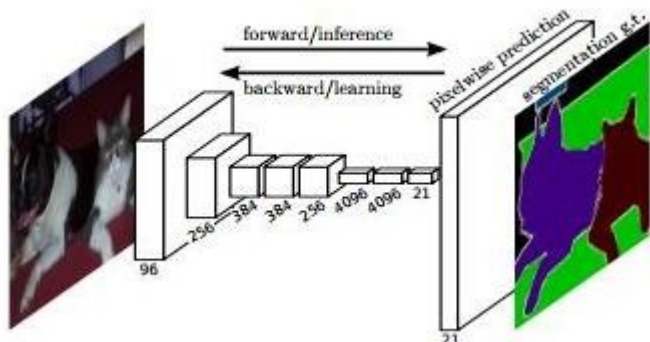
выход — карты поточечной оценки

для каждого класса своя карта

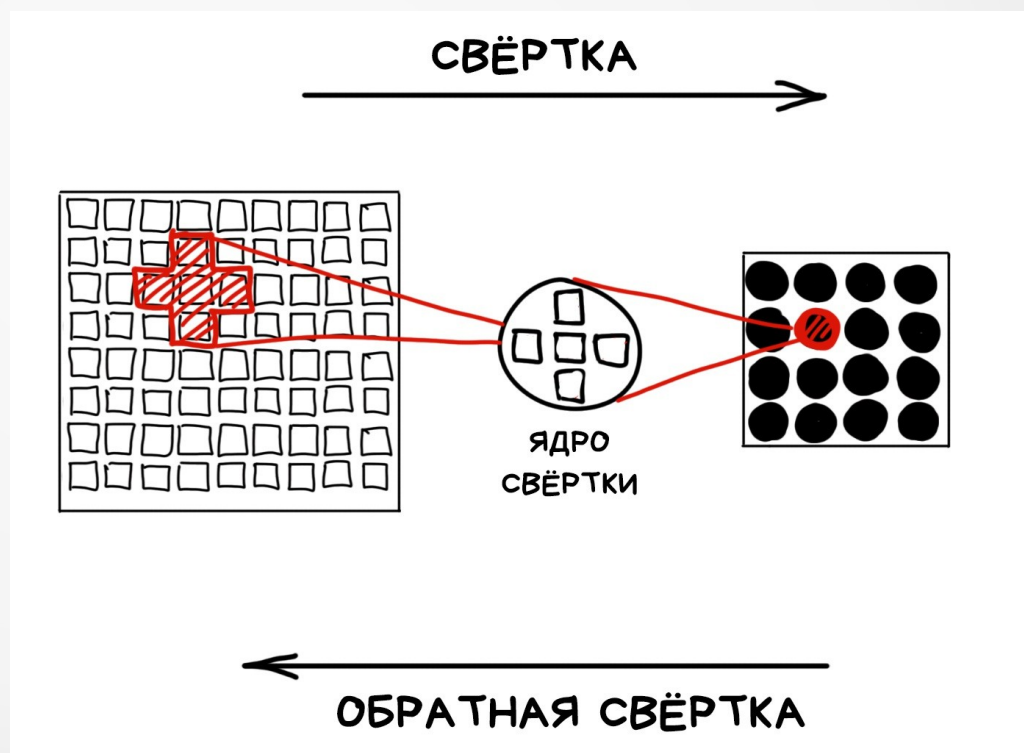
размер входного изображения = размеру входной карты

сравниваем выходные карты поточечно,  
для каждой точки определяем карту-победителя

# Поиск объектов на изображении



изображение обрабатывается свёрточными слоями  
на выходе выполняем обратную свёртку



# Поиск объектов на изображении

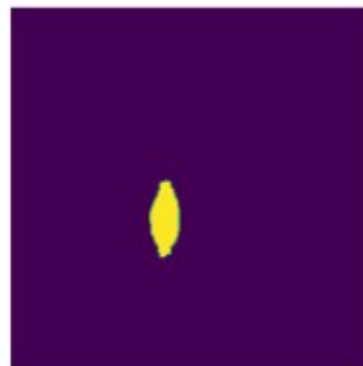
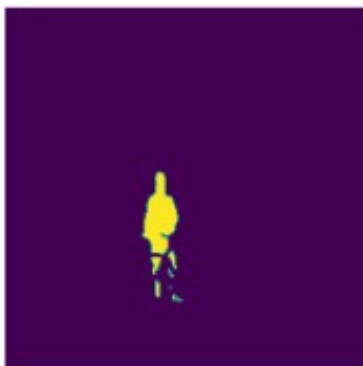
## FCN: Fully Convolutional Networks

Пример — ищем людей на картинке (датасет Pascal VOC)

картинка

разметка

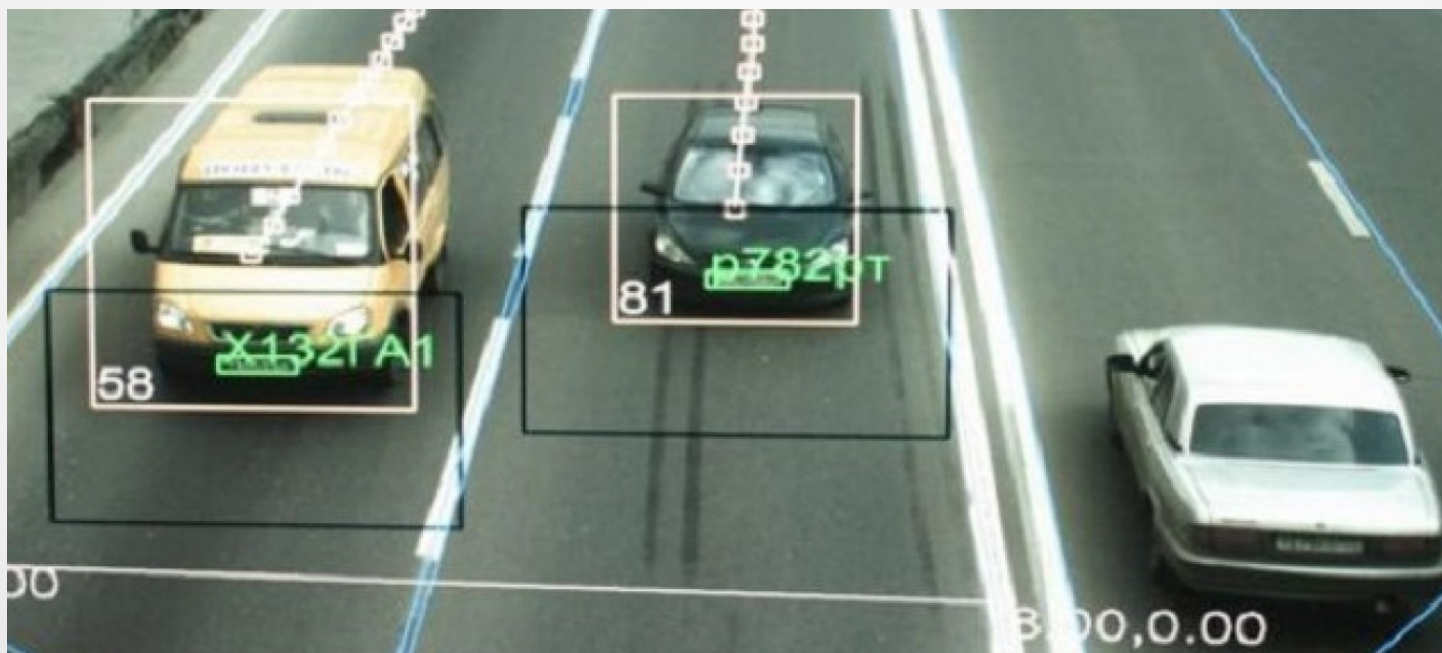
результат FCN



# Поиск объектов на изображении

## Модель объекта с помощью нейросети

Задача классификации

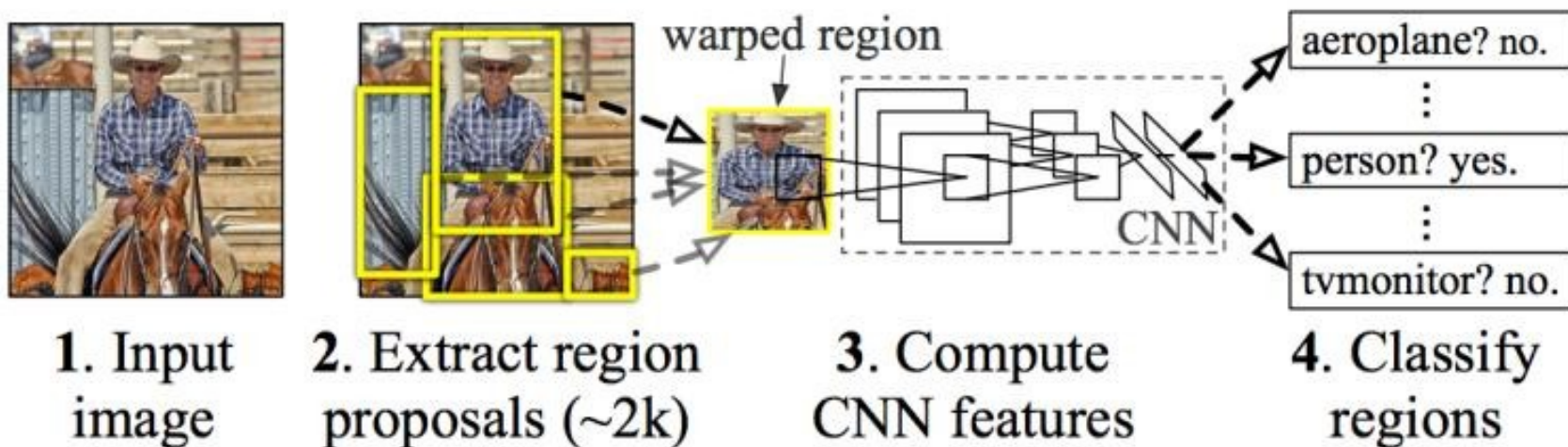




# Поиск объектов на изображениях

## Object detection

### R-CNN: *Regions with CNN features*



### Region Based Convolutional Neural Networks (R-CNN)

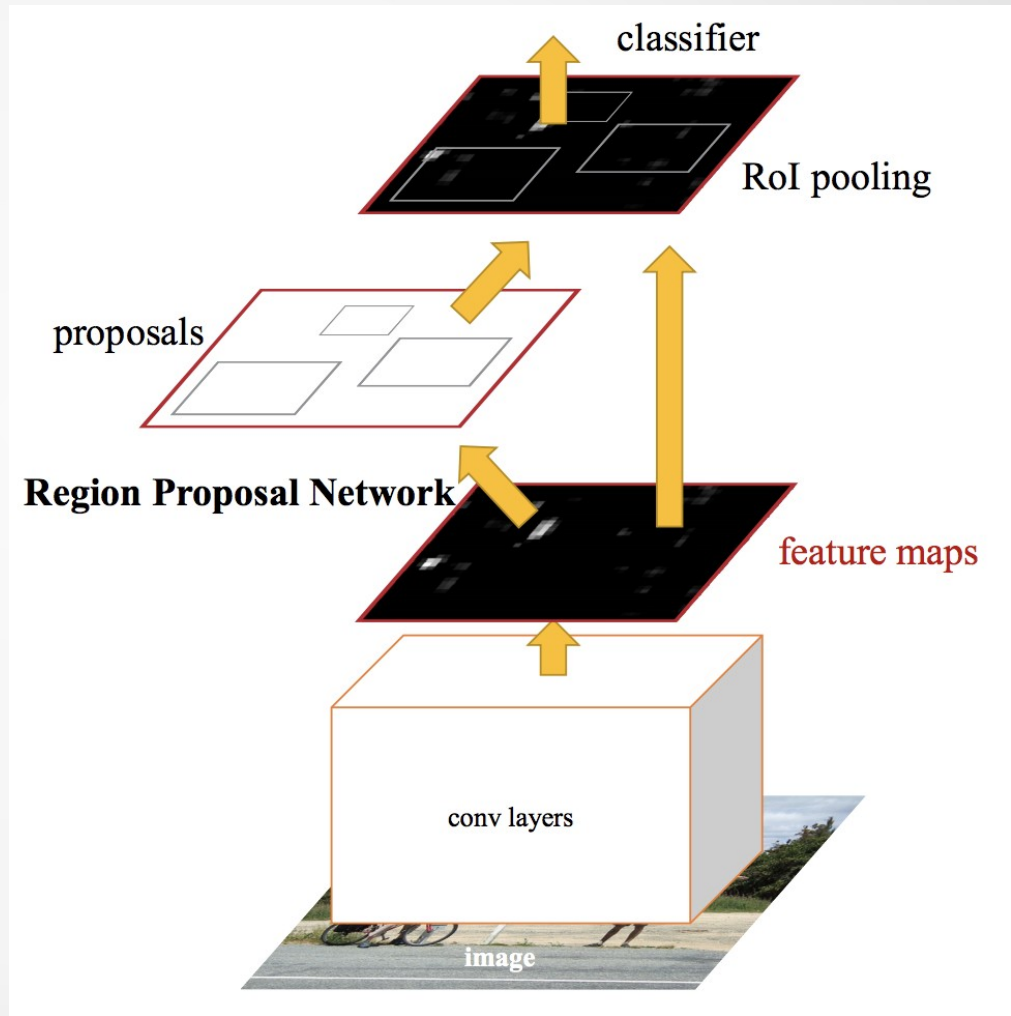
изображение разделяется на части

каждую часть проверяем классификатором

# Поиск объектов на изображении

## Faster-R-CNN

- принимаем картинку на вход
- картинка прогоняется через CNN, формируем feature maps
- определяем регионы-кандидаты (возможно содержащие объекты)
- выделяем эти регионы
- и применяем к ним классификатор картинок



# Свёрточные нейросети

## **Датасеты**

MNIST (National Institute of Standards and Technology)

CIFAR-10 (Canadian Institute for Advanced Research)

Pascal VOC (Visual Object Classes )

ImageNet

# Свёрточные нейросети

## MNIST (National Institute of Standards and Technology)

28x28 grayscale, 60K training images, 10K testing images, 10 classes

<https://github.com/zalandoresearch/fashion-mnist>

[https://ru.wikipedia.org/wiki/MNIST\\_\(база\\_данных\)](https://ru.wikipedia.org/wiki/MNIST_(база_данных))

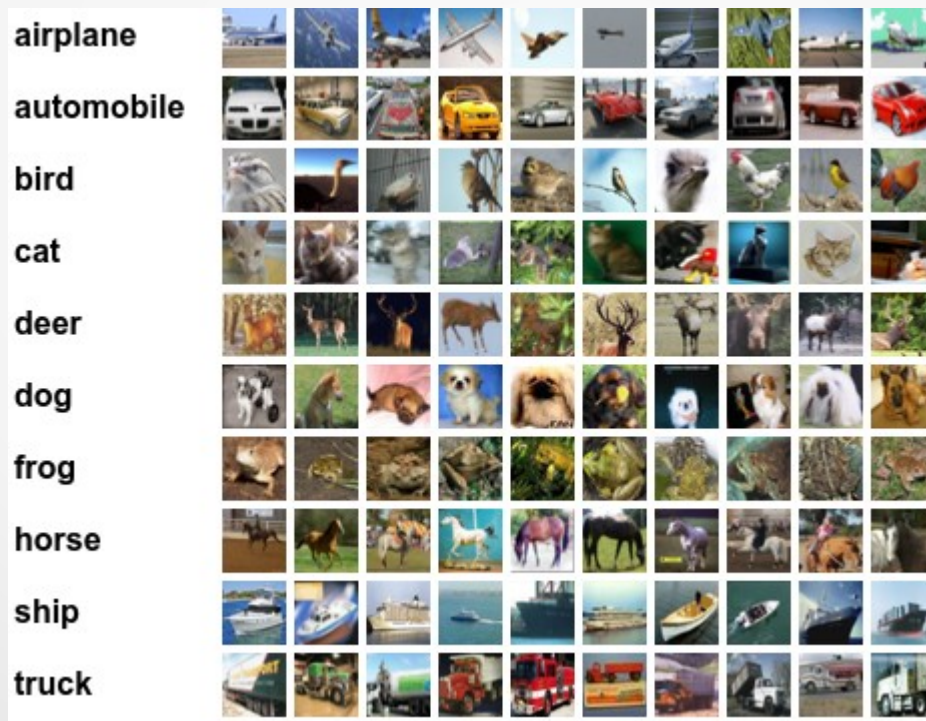




# Свёрточные нейросети

**CIFAR-10 (Canadian Institute for Advanced Research)**  
32x32 color, 60K images, 10 classes

<https://www.cs.toronto.edu/~kriz/cifar.html>



# Свёрточные нейросети

## ImageNet

14М изображений, 21К категорий

<http://www.image-net.org>



ILSVRC (ImageNet Large Scale Visual Recognition Challenge)



# Свёрточные нейросети

## **Pascal VOC (Visual Object Classes )**

11K images, 20 classes, 27K ROI annotated objects and 7K segmentations

<http://host.robots.ox.ac.uk/pascal/VOC/pubs/everingham10.pdf>



# Поиск объектов на изображении

## Литература

Борисов Е.С. Методы машинного обучения. 2024  
[https://github.com/mechanoid5/ml\\_lectorium\\_2024\\_I](https://github.com/mechanoid5/ml_lectorium_2024_I)

Борисов Е.С. Базовые методы обработки изображений.  
<http://mechanoid.su/cv-base.html>

Борисов Е.С. О задаче поиска объекта на изображении.  
<http://mechanoid.su/cv-image-detector.html>

Конущин А.С. Введение в компьютерное зрение. 2015  
[https://www.youtube.com/playlist?list=PL-\\_cKNuVAYAXAnpy8RCV8UtFrFFLRa4rh](https://www.youtube.com/playlist?list=PL-_cKNuVAYAXAnpy8RCV8UtFrFFLRa4rh)

Николенко С., Кадури́н А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей. - "Питер", 2018 г.

Fully Convolutional Networks for Semantic Segmentation  
<https://arxiv.org/pdf/1411.4038.pdf>

FCN — Fully Convolutional Network (Semantic Segmentation)  
<https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1>