

Занятие № 9

Свёрточные нейронные сети. Часть 1

План занятия



1. Что такое изображения и какие они бывают
2. Задачи обработки изображений
3. Классические алгоритмы обработки изображений
4. Детектирование объектов на изображении
5. Сверточные слои
6. Операции пулинга
7. Архитектура LeNet
8. Архитектура VGG

Изображения

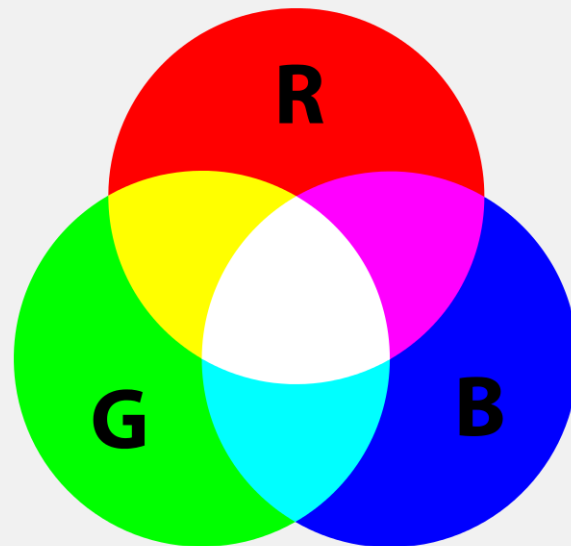


Параметры изображения:

Размеры: (w, h)

Каналы: c

Динамический диапазон: d



Типичные изображения:

Цветные (RGB) - $(w, h, 3)$, 8 bit

Черно-белые (Grayscale) - (w, h) или $(w, h, 1)$, 8 bit

Задачи обработки изображений

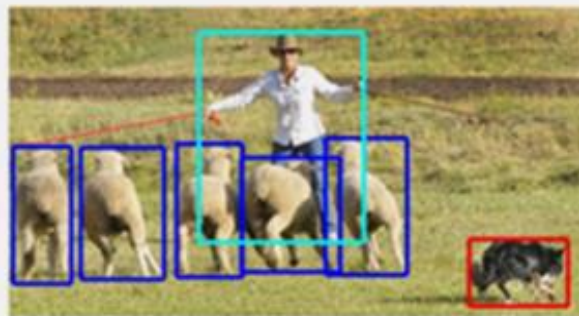


Основные задачи обработки изображений:

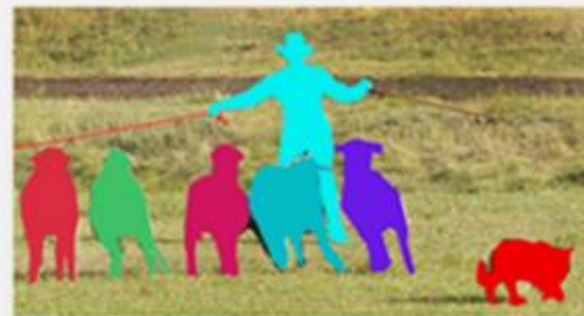
- a. Классификация объектов
- b. Детектирование объектов
- c. Сегментация объектов



(a) classification



(b) detection



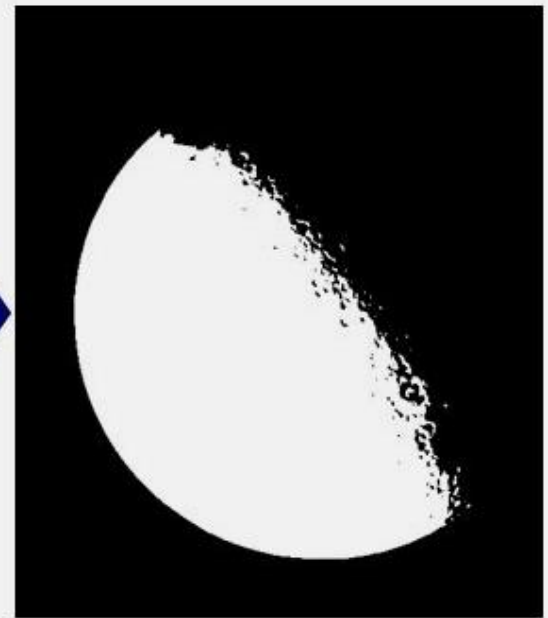
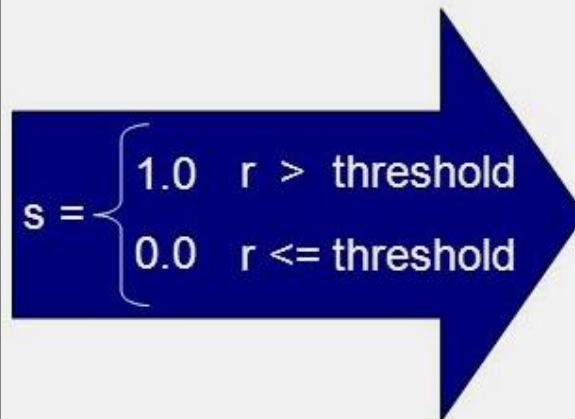
(c) segmentation

Классическая сегментация по порогу



Пусть дано 1-канальное изображение.

Пиксели, где яркость выше определенного значения считаем объектом, остальное фон



Классическая сегментация по порогу



Порог может быть:

- Константой для набора изображений (задаем единый порог для всех изображений)
- Константой для одного изображения набора (по какому-либо правилу вычисляем порог для целого изображения)
- Адаптивный (по какому-либо правилу вычисляем порог для каждого пикселя изображения)

Адаптивный порог



What Is Image Filtering in the Spatial Domain?

Filtering is a technique for modifying or enhancing an image. For example, you can filter an image to emphasize certain features or remove other features. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement.

Filtering is a *neighborhood operation*, in which the value of any given pixel in the output image is determined by applying some algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. A pixel's neighborhood is some set of pixels, defined by their locations relative to that pixel. (See Neighborhood or Block Processing: An Overview for a general discussion of neighborhood operations.) *Linear filtering* is filtering in which the value of an output pixel is a linear combination of the values of the pixels in the input pixel's neighborhood.

Convolution

Linear filtering of an image is accomplished through an operation called *convolution*. Convolution is a neighborhood operation in which each output pixel is the weighted sum of neighboring input pixels. The matrix of weights is called the *convolution kernel*, also known as the *filter*. A convolution kernel is a correlation kernel that has been rotated 180 degrees.

For example, suppose the image is

```
A = [17 24 1 8 15
      23 5 7 14 16
      4 6 13 20 22
      10 12 19 21 3]
```

Global threshold

What Is Image Filtering in the Spatial Domain?

Filtering is a technique for modifying or enhancing an image. For example, you can filter an image to emphasize certain features or remove other features. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement.

Filtering is a *neighborhood operation*, in which the value of any given pixel in the output image is determined by applying some algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. A pixel's neighborhood is some set of pixels, defined by their locations relative to that pixel. (See Neighborhood or Block Processing: An Overview for a general discussion of neighborhood operations.) *Linear filtering* is filtering in which the value of an output pixel is a linear combination of the values of the pixels in the input pixel's neighborhood.

Linear filtering of an image is accomplished through an operation called *convolution*. Convolution is a neighborhood operation in which each output pixel is the weighted sum of neighboring input pixels. The matrix of weights is called the *convolution kernel*, also known as the *filter*. A convolution kernel is a correlation kernel that has been rotated 180 degrees.

Adaptive threshold

What Is Image Filtering in the Spatial Domain?

Filtering is a technique for modifying or enhancing an image. For example, you can filter an image to emphasize certain features or remove other features. Image processing operations implemented with filtering include smoothing, sharpening, and edge enhancement.

Filtering is a *neighborhood operation*, in which the value of any given pixel in the output image is determined by applying some algorithm to the values of the pixels in the neighborhood of the corresponding input pixel. A pixel's neighborhood is some set of pixels, defined by their locations relative to that pixel. (See Neighborhood or Block Processing: An Overview for a general discussion of neighborhood operations.) *Linear filtering* is filtering in which the value of an output pixel is a linear combination of the values of the pixels in the input pixel's neighborhood.

Convolution

Linear filtering of an image is accomplished through an operation called *convolution*. Convolution is a neighborhood operation in which each output pixel is the weighted sum of neighboring input pixels. The matrix of weights is called the *convolution kernel*, also known as the *filter*. A convolution kernel is a correlation kernel that has been rotated 180 degrees.

For example, suppose the image is

```
A = [17 24 1 8 15
      23 5 7 14 16
      4 6 13 20 22
      10 12 19 21 3]
```

Классификация сегментированных объектов, свойства формы



- Площадь
- Относительная протяженность границ
- Радиусы вписанной и описанной окружности
- И другие

Сведение задачи к сегментации по порогу

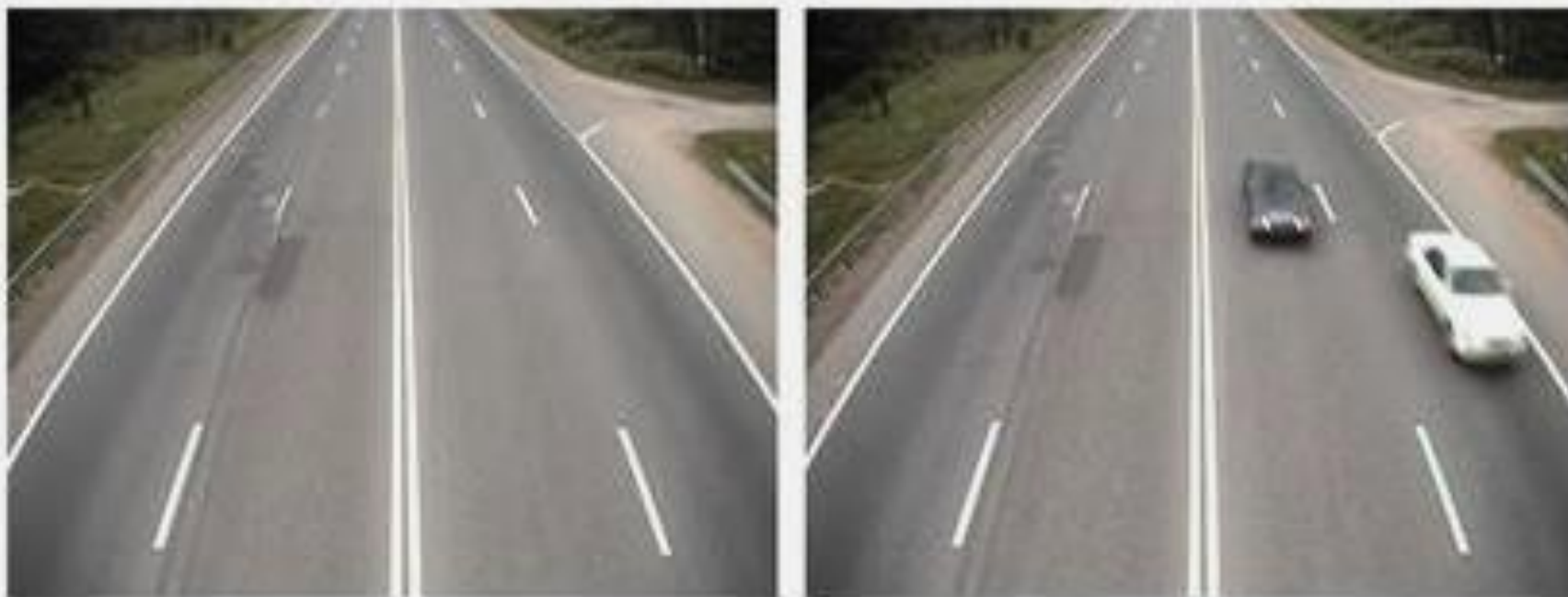


- Преобразование цветового пространства
- Вычитание фона
- Фильтрация

Вычитание фона



Вычитание фона позволяет эффективно сегментировать объекты при условии статичности фона



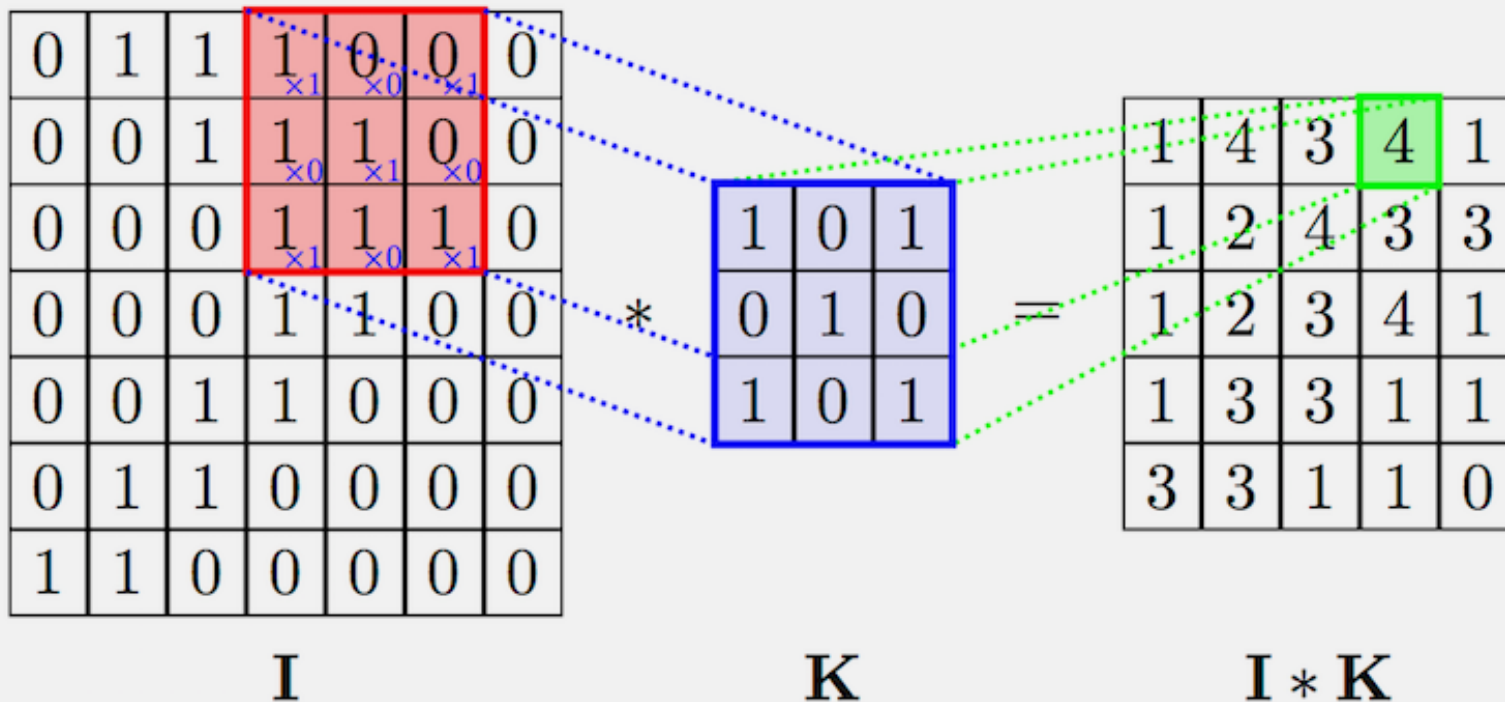
Фильтрация, операция свертки



Дано:

Исходное изображение **I**, массив (w, h)

Ядро свертки **K**, массив (x, y), $x < w$, $y < h$



Поиск границ, оператор Собеля (пример фильтрации)



Идея: посчитаем градиент яркости изображения

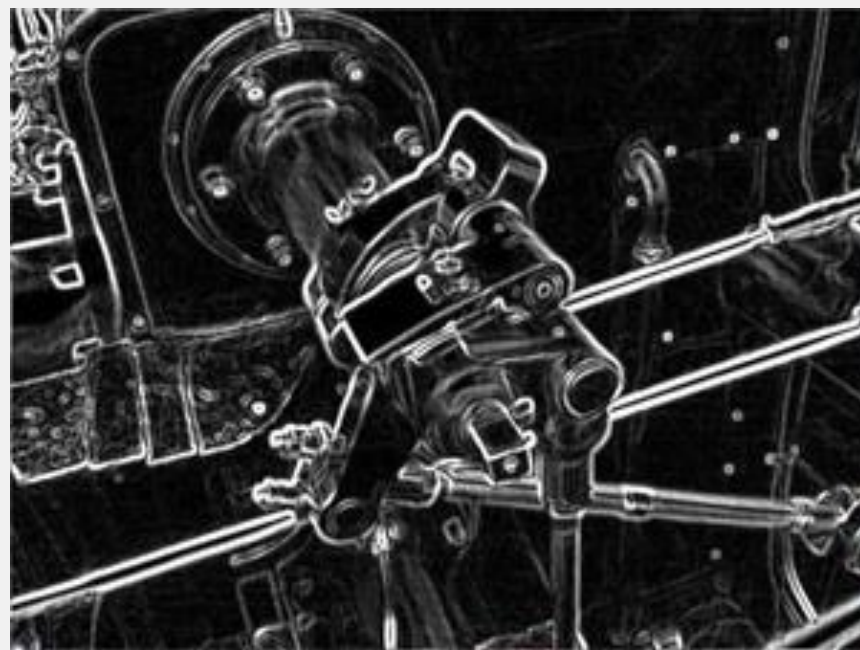
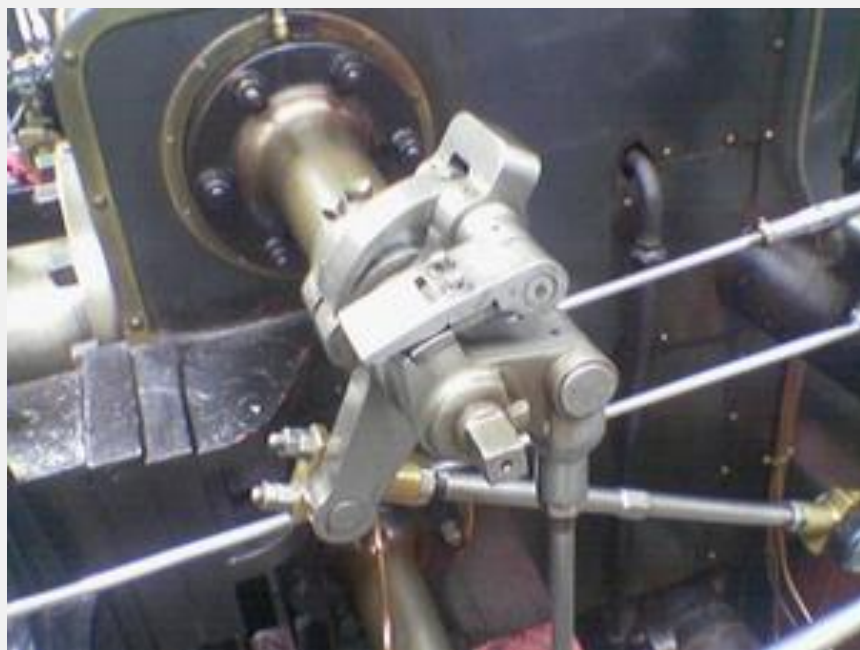
$$\mathbf{G}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * \mathbf{A} \quad \text{and} \quad \mathbf{G}_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * \mathbf{A}$$

$$\mathbf{G} = \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2}$$

Оператор Собеля, поиск границ



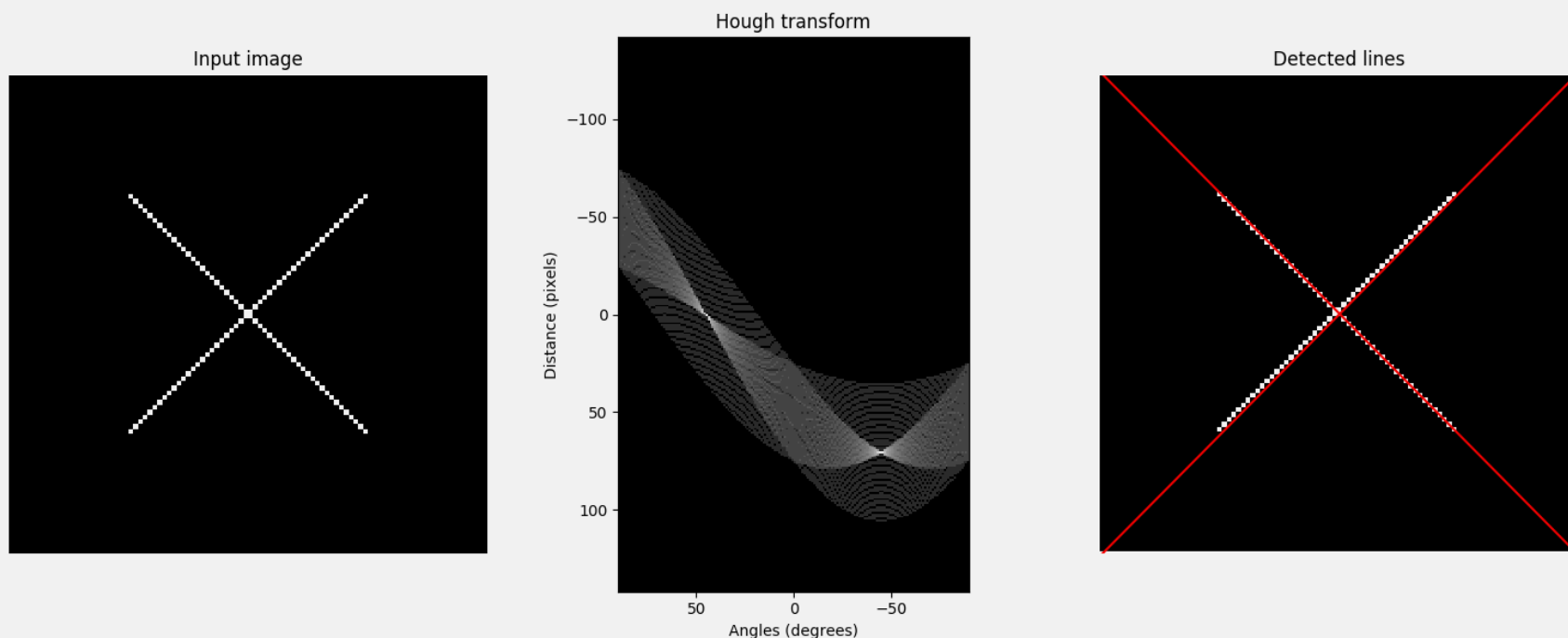
Пример работы оператора Собеля



Работа с границами, преобразование Хафа



Преобразование Хафа используется для поиска объектов, принадлежащих определённому классу фигур, с использованием процедуры голосования.



Template Matching

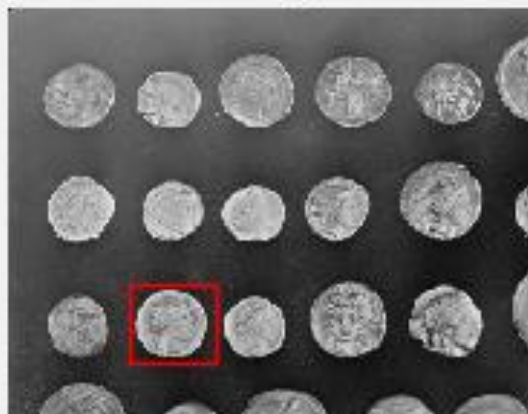


Проходим по изображению сравнивая эталонное изображение с фрагментами исследуемого каким-либо образом

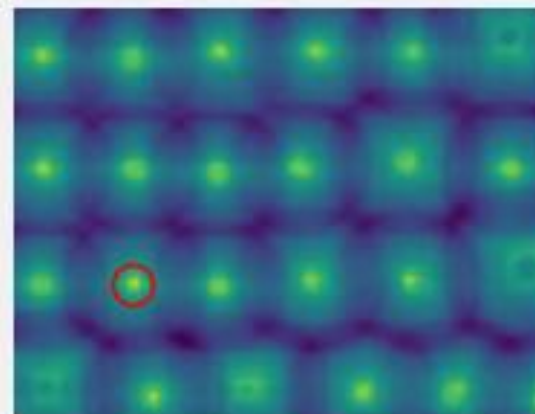
template



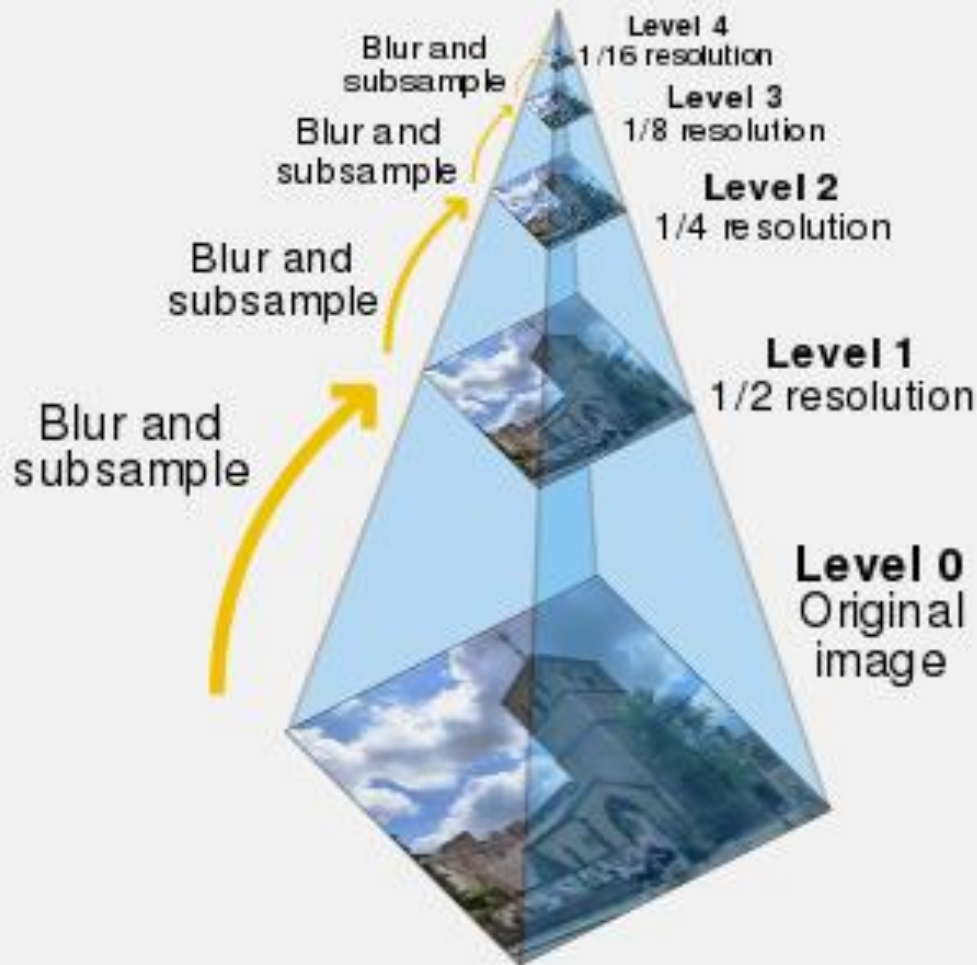
image



`match_template`
result



Проблема масштаба, пирамиды изображений



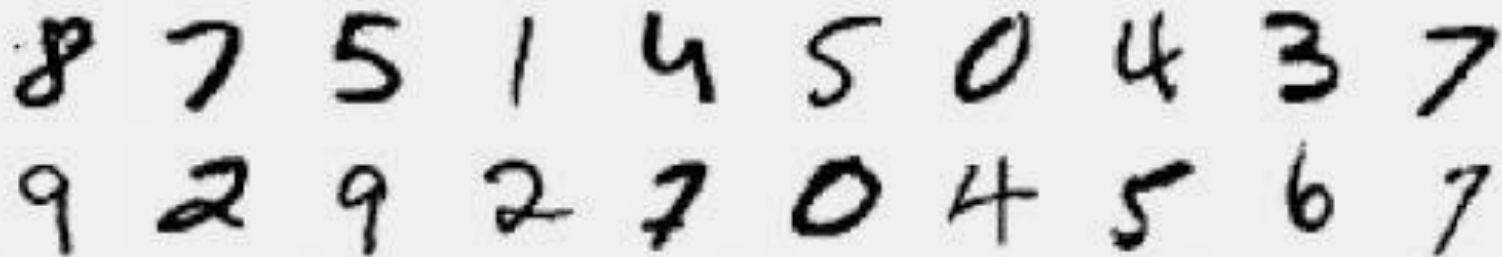
Пирамиды изображений позволяют выполнять Template Matching в различных масштабах

Машинное обучение для классификации фрагментов изображений



Яркости пикселей небольшого изображения можно напрямую использовать как признаки и обучить на них модель машинного обучения

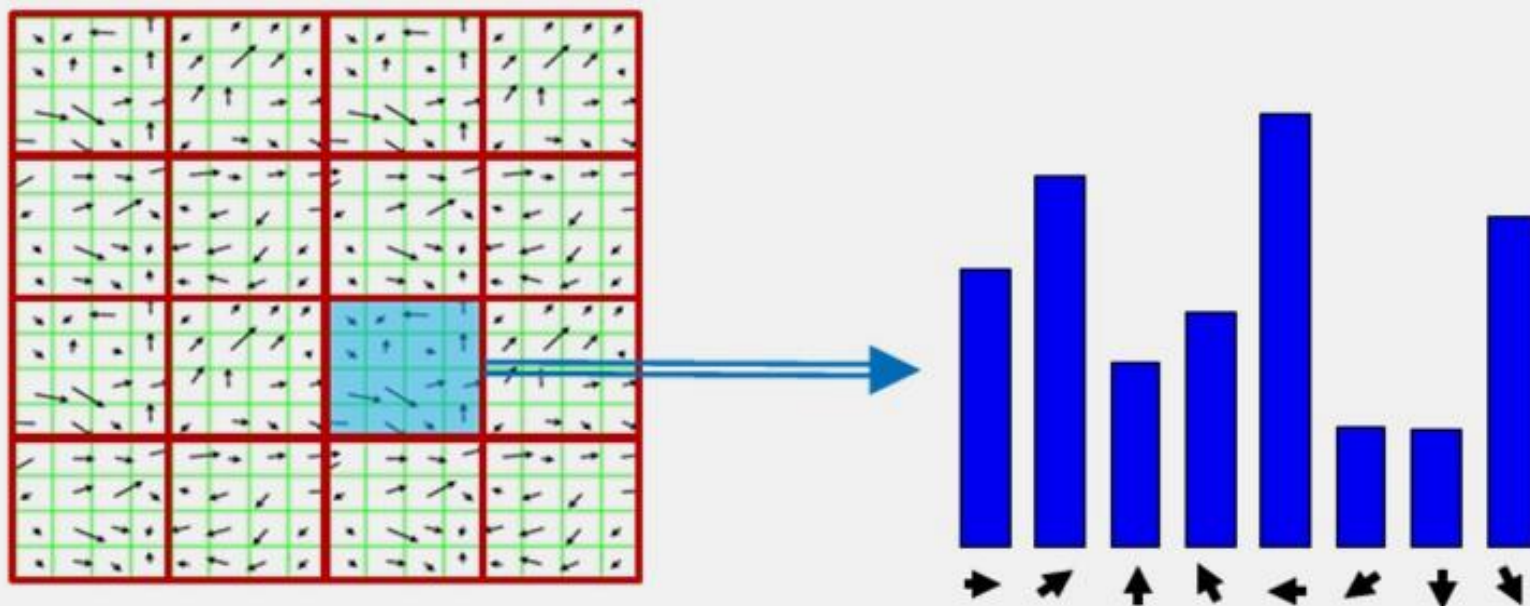
На наборе изображений рукописных цифр MNIST можно таким способом получить точность 98%



Дескриптор HOG



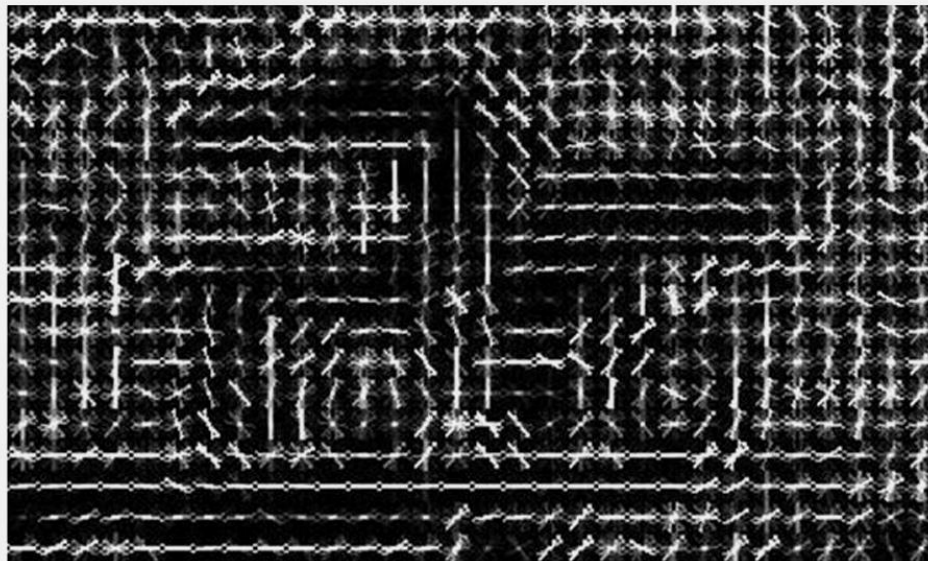
Гистограмма направленных градиентов (HOG, Histogram of Oriented Gradients) - метод представления изображения в виде вектора признаков



Дескриптор HOG



Пример:

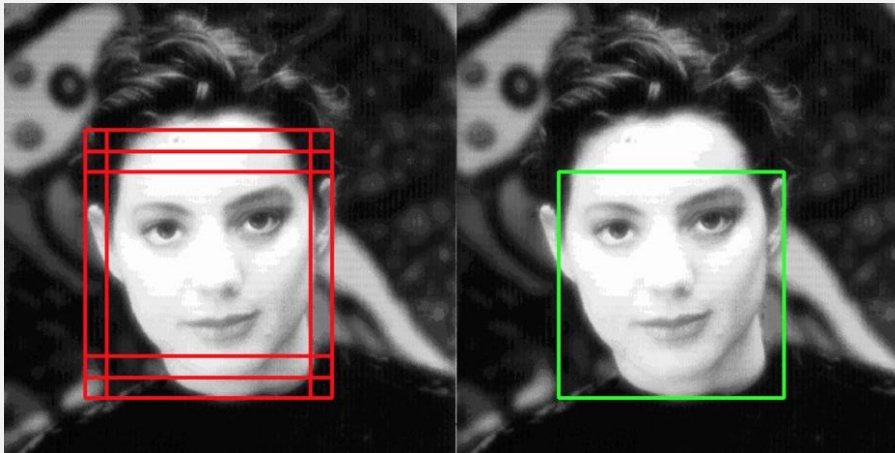


Детектор объектов



Пирамиды картинок + алгоритм машинного обучения на фрагментах = детектор объектов

Проблема:



Решение: non-maximum suppression

Современные подходы



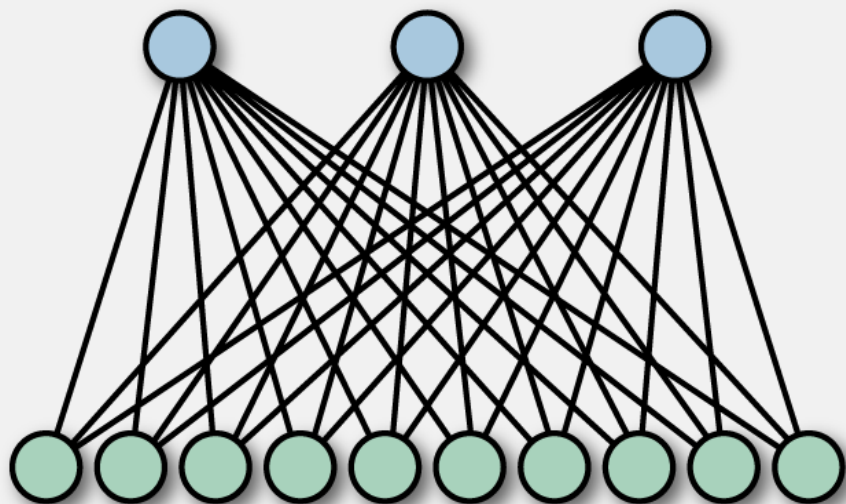
Современные методы анализа изображений основаны на **сверточных нейронных сетях**

Основное отличие **современных** подходов от **классических** заключается в том, что современные подходы используют **автоматическую** генерацию признаков из изображений

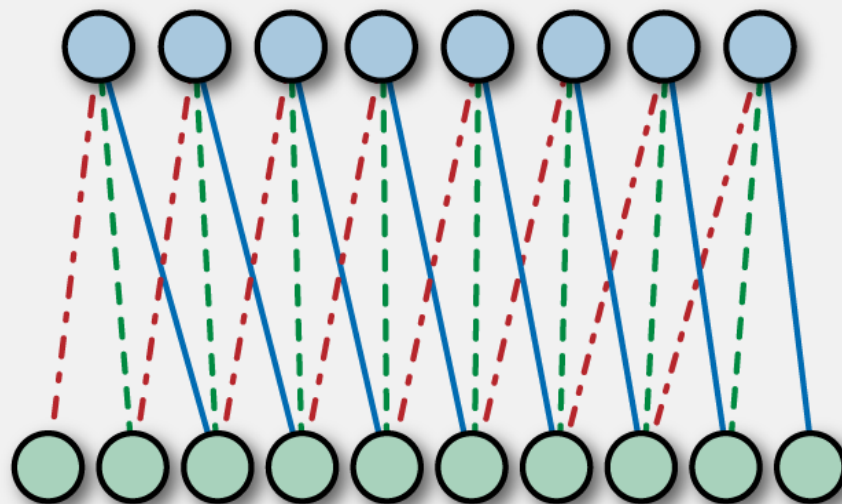
Сверточные слои



Fully Connected



Convolutional Layer



Сверточный слой - это полносвязный слой в котором мы игнорируем не локальные связи, а локальные связи считаем общими для всех нейронов

Сверточные слои



Для входа 100×100 и выхода 100×100 :

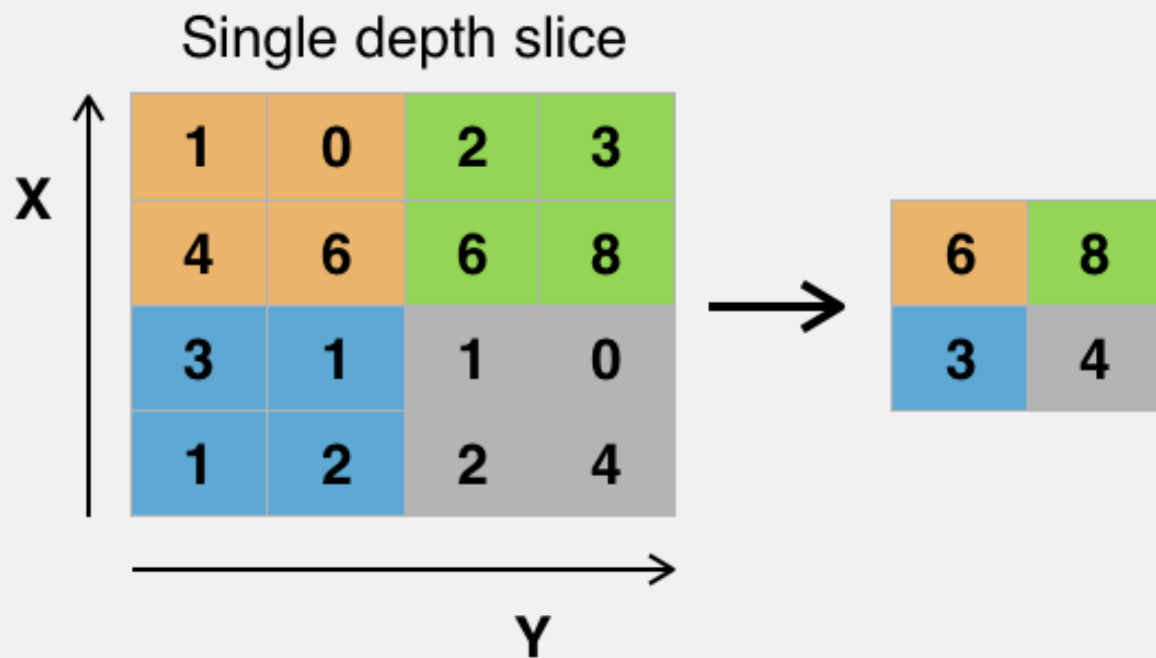
- **Полносвязный слой** имеет 10^8 параметров.
- **Сверточный слой** с ядром $K 5 \times 5$ имеет 25 параметров и не зависит от размера входного изображения!

Слой пулинга



Слой **пулинга** (pooling) - это слой снижения размерности карт признаков

Для снижения размерности как правило используют функцию **max**





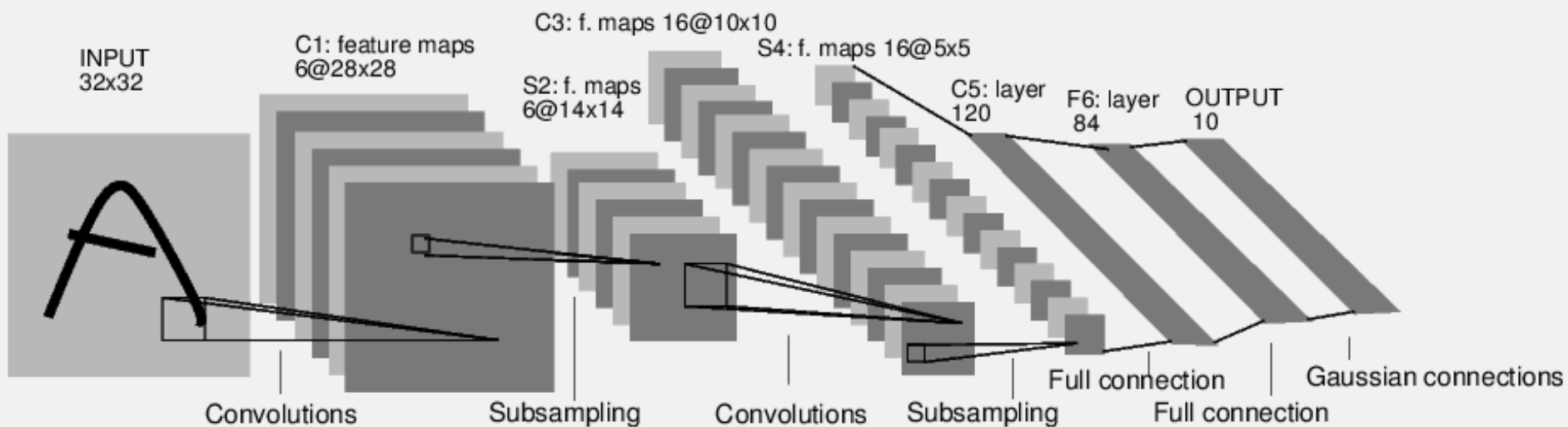
Основные параметры сети:

- Перед подачей в сеть изображения стандартизировали
- Свертки 5x5
- Пулинг с функцией sum
- Сигмоидальная функция активации

(LeCun et al., 1998)



Архитектура сети:



Что такое ImageNet?



ImageNet - проект, направленный на (ручную) маркировку и категоризацию изображений в примерно 22 000 отдельных категорий объектов для задач исследований в области компьютерного зрения.

В рамках проекта **ImageNet** ежегодно проводится соревнование ImageNet Large Scale Visual Recognition Challenge (**ILSVRC**)



Задача соревнования - построить модель, для классификации изображений в 1000 отдельных категорий объектов. Эти 1000 категорий изображений представляют классы предметов, с которыми мы сталкиваемся в повседневной жизни, такие как виды собак, кошек, различные предметы быта, типы транспортных средств и многое другое.

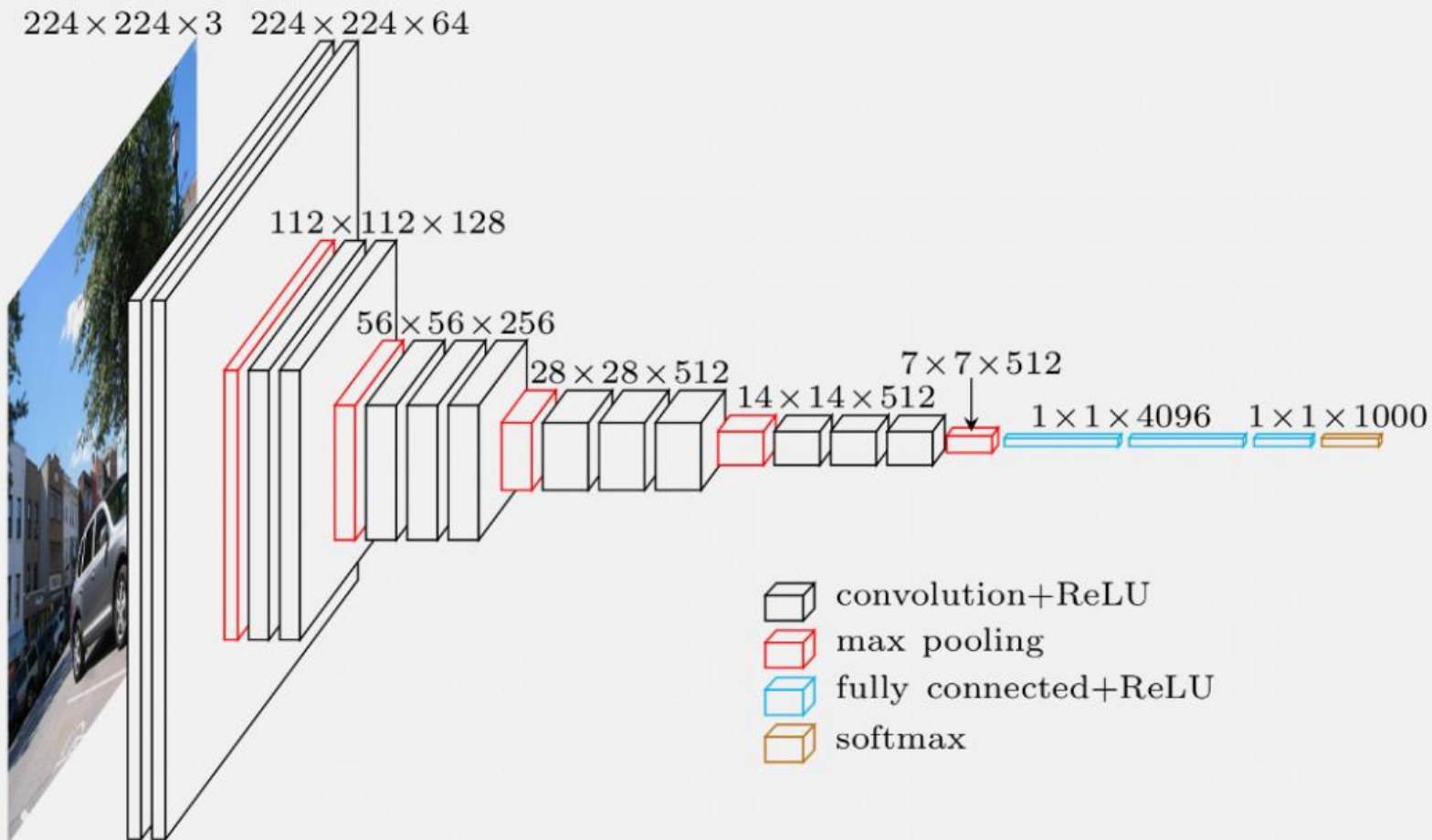
Train: ~1.2 миллиона изображений

Test: ~100 тысяч изображений

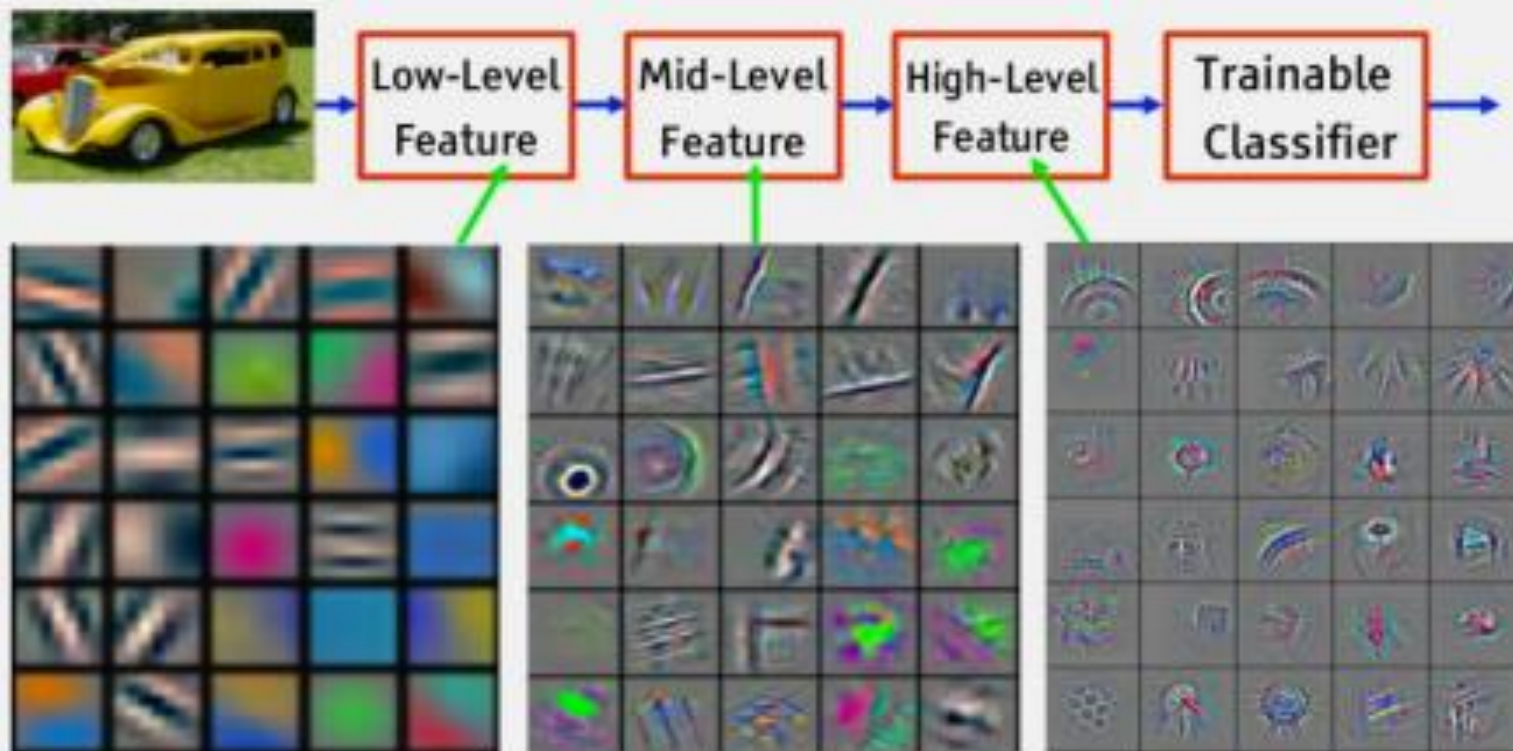
VGG-16



Архитектура сети VGG-16 (победитель ILSVRC 2014)



Что выучивают нейронные сети?





**Спасибо за
внимание!**

Евгений Некрасов

e.nekrasov@corp.mail.ru