Выделение объектов и глубокие нейронные сети

Александр Сергеев

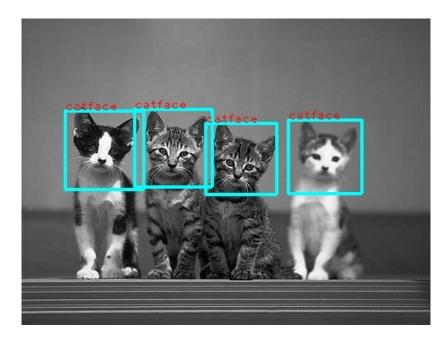
НИУ Высшая школа экономики

26 мая 2016 г.

Описание задачи

Bходные данные Изображение I или видео $S = \{I_k\}, \ k = 1, 2, 3, \ldots$

Выходные данные $\{R_i\},\ i=\overline{1,n},\$ где R_i — ограничивающий прямоугольник, n — число обнаружений, или $\{Out_k\,|\,I_k\in S\}\ Out_k=\{R_i\},\ i=\overline{1,n_k},\$ где n_k — число выделенных объектов на k-том кадре.



Задачи распознавания

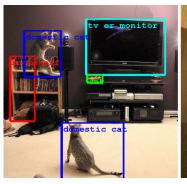
Классификация изображений (ImageNET):

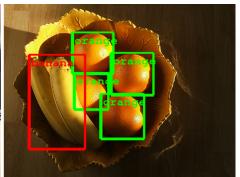


14 миллионов изображений, 20 тысяч классов.

Задачи распознавания

Выделение (локализация* объектов):





Задачи распознавания

Сегментация представленной сцены:



Детекторы объектов

Ранняя классификация:

- knowledge-based;
- appearance-based.

Сейчас выделить три основных направления:

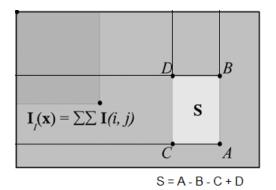
- на основе интегральных каналов (ACF/ICF);
- деформируемые модели частей (DPM);
- глубокие сверточные нейронные сети (CNN).

Детектор лиц Виола-Джонса:

- фильтры на прямоугольных областям;
- бустинг слабых признаков (AdaBoost);
- каскад бинарных классификаторов.

Выделение объектов сводится к бинарной классификации.

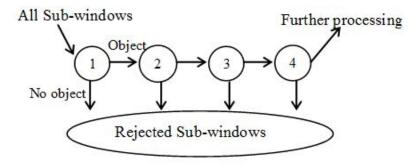
Как классификатор используется неглубокое дерево решений на основе откликов фильтров на интегральном изображении.



AdaBoost:

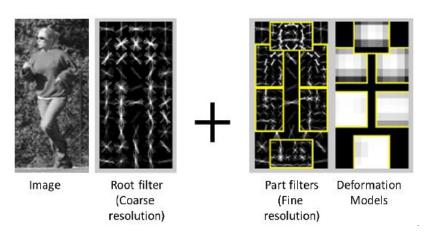
- строит линейный классификатор из «слабых» классификаторов;
- последовательно добавляет новый (лучший) классификатор;
- использует веса для примеров из обучающей выборки.

Каскад позволяет уменьшить объем вычислений:



Deformable Part Model

Поиск объектов с подвижной моделью частей.



В качестве признаков используется HoG.

Deformable Part Model

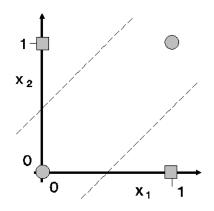
Используется скользящее окно, при этом для каждого кандидата:

$$score(p_0, ..., p_n) = \sum_{i=0}^{n} A_i - \sum_{i=1}^{n} d_i \cdot def(\Delta x_i, \Delta y_i) + b$$
$$def(\Delta x, \Delta y) = (\Delta x, \Delta y, \Delta x^2, \Delta y^2)$$

 d_i — настраиваемые веса, например, (0,0,1,1). Для выбора наиболее подходящего представления используется выбор максимума.

1960-1970: перцептрон, однослойные сети:

$$f(x) = \begin{cases} 1, \sum_{i=1}^{n} w_i x_i + b \ge 0\\ 0, \text{ otherwise} \end{cases}$$



1970-1990: многослойные сети, выбор функций активации, метод обратного распространения ошибки.

1998: Yann LeCun: описывает сверточные сети для распознавания рукописных цифр

2009-2012: создание ImageNET, высокие показатели deep cnn.

Стохастический градиентный спуск (SGD):

$$w_i \leftarrow w_i - \eta \frac{\partial E}{\partial w_i},$$

где E — функция потерь, w_i — параметры сети.

SGD с использованием момента (momentum):

$$v \leftarrow \gamma v + \eta \frac{\partial E}{\partial w_i}$$
$$w_i \leftarrow w_i - v$$

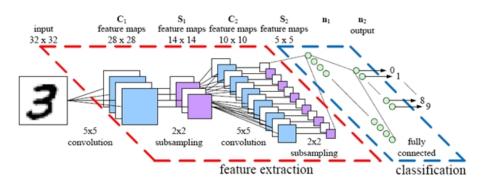
SGD:

$$w_i \leftarrow w_i - \eta \frac{\partial E}{\partial w_i}$$

Регуляризация (затухание весов):

$$E'(w) = E(w) + \frac{\lambda}{2}w^2$$
$$w_i \leftarrow w_i - \eta \frac{\partial E}{\partial w_i} - \eta \lambda w_i$$

Сверточная нейронная сеть



Проблемы DCNN

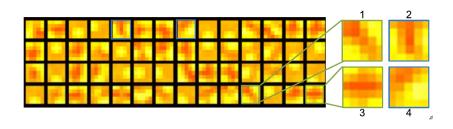
- vanishing/exploding gradients;
- размер данных и выбор модели;
- инициализация весов;
- выбор функции потерь.

Движение к выделению объектов:

- классификация изображений, «богатые» признаки;
- внешняя генерация гипотез и классификатор;
- каскадный нейросетевой классификатор;
- нейросетевой генератор гипотез (RPN);

Def-pooling layer

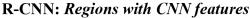
DPM можно представить в виде CNN.

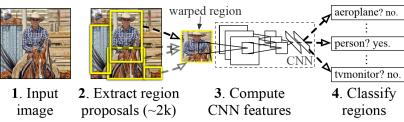


$$F_p^{def} = \max_{d_i, d_i \in \{-R, \dots, R\}} \{F_p^{part}[p_i + d_i, p_j + d_j] - a_p d_p^{d_i, d_j}\}$$

То есть, обучаем матрицы штрафов и их веса.

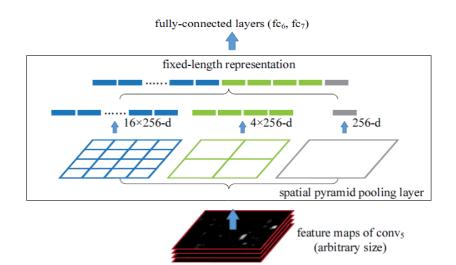
Region-based CNN



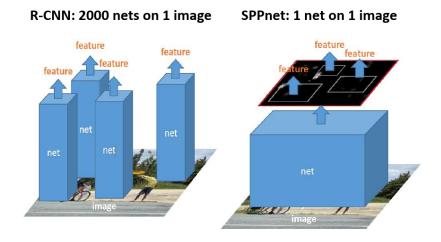


Генерация гипотез через внешний алгоритм: selective search, edgeboxes.

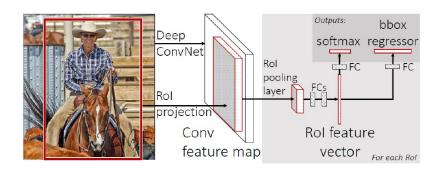
Spartial Pyramid Pooling



Spartial Pyramid Pooling

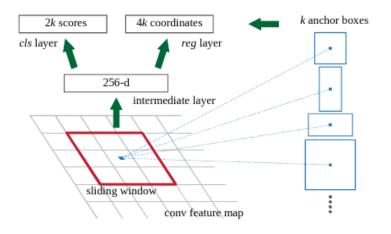


Fast R-CNN



Используется внешний генератор гипотез.

Faster R-CNN



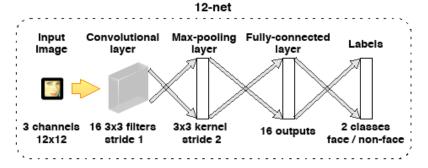
Совместное обучение классификатора и генератора гипотез (RPN).

Каскадные нейросети

Каскад из последовательности нейронных сетей парами: бинарный классификатор и калибрующая сеть (изображения 12х12, 24х24, 48х48)

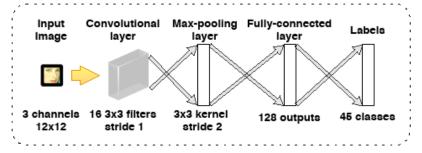
Размер и сложность сетей возрастает; классификаторы отбрасывают часть примеров.

Классификатор



Калибратор

12-calibration-net



Отдельные интересные темы:

- residual learning, ResNet, FractalNet;
- adversarial examples;
- weakly supervised learning;
- data-dependent initialization;
- difficulty control, self-paced learning.

Разрабатываемые темы

- представление каскада из слоев сети с bbox как данными;
- использование глобального контекста сцены для управления свертками;
- неполная разметка для выделения людей на камерах;
- развитие использования непосредственного контекста.