О задаче поиска объектов на изображении. Свёрточные сети.

Евгений Борисов

Датасеты

MNIST (National Institute of Standards and Technology)

CIFAR-10 (Canadian Institute for Advanced Research)

Pascal VOC (Visual Object Classes)

ImageNet

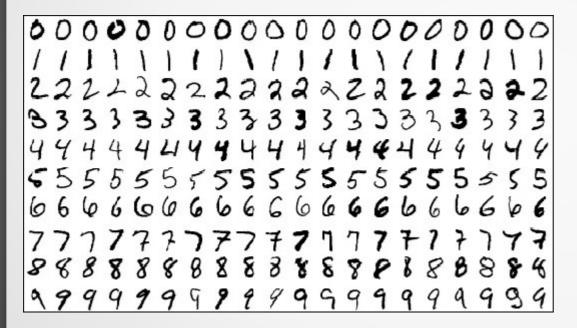
MS COCO (Microsoft Common Objects in Context)

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)













lbl: Ankle boot



pred: Ankle boot

lbl: Sandal

 $\alpha \pi$

pred: Sandal

Ibl: Coat

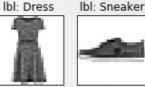


lbl: Sandal

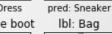
pred: Sandal







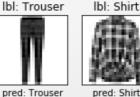














pred: Bag



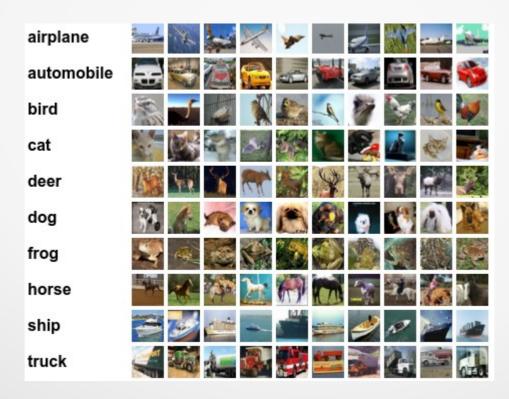


pred: Bag

CIFAR-10 (Canadian Institute for Advanced Research)

32x32 color, 60K images, 10 classes

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



ImageNet

14М изображений, 21К категорий



http://www.image-net.org

ILSVRC (ImageNet Large Scale Visual Recognition Challenge)



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



MS COCO (Microsoft Common Objects in Context)

328 000 изображений

https://cocodataset.org



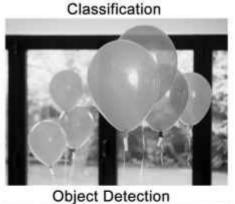
Задача поиска объектов на изображении

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

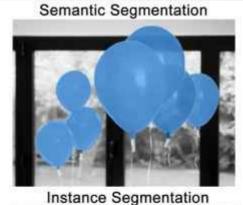
Semantic segmentation — определение всех пикселей объектов определённого класса или фона на изображении. Если несколько объектов одного класса перекрываются, их пиксели никак не отделяются друг от друга;

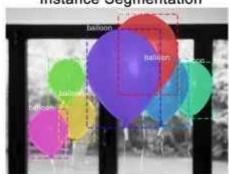
Object detection — обнаружение всех объектов указанных классов и определение охватывающей рамки для каждого из них;

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



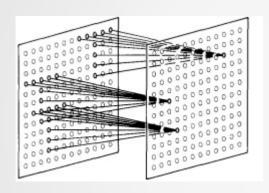




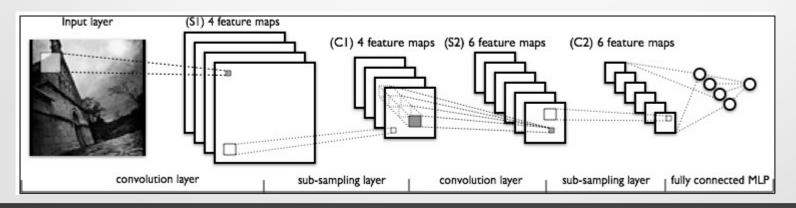


Свёрточные сети

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.

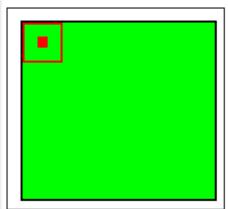


Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, Winter 1989.



Операция свёртки

$$(f*g)[m,n] = \sum_{k,l} f[m-k,n-l] \cdot g[k,l]$$





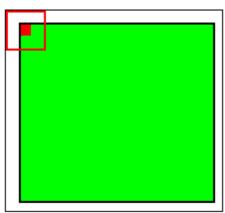


Рис.3: обработка краёв same

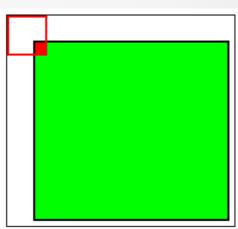


Рис.4: обработка краёв full

g -ядро свёртки

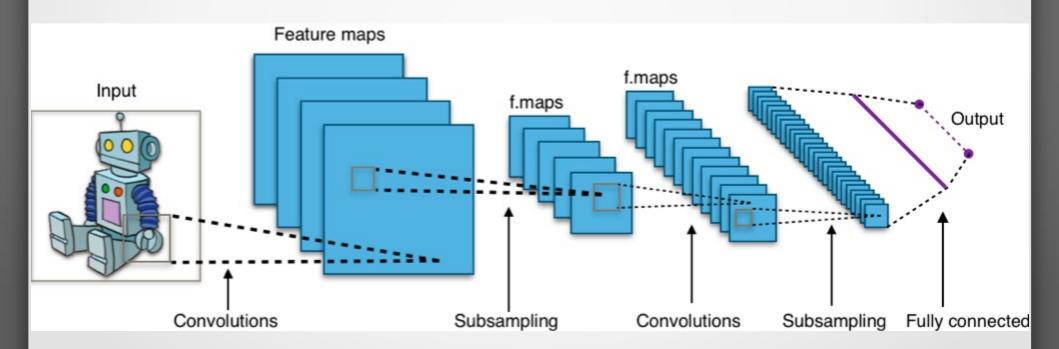
- берём точку с окрестностью,
- поэлементно умножаем эту матрицу на ядро, результат суммируется и записывается как новое значение данной точки
- процедура повторяется для всех точек изображения.

Свёрточная сеть LeNet

свёрточный слой (convolution)

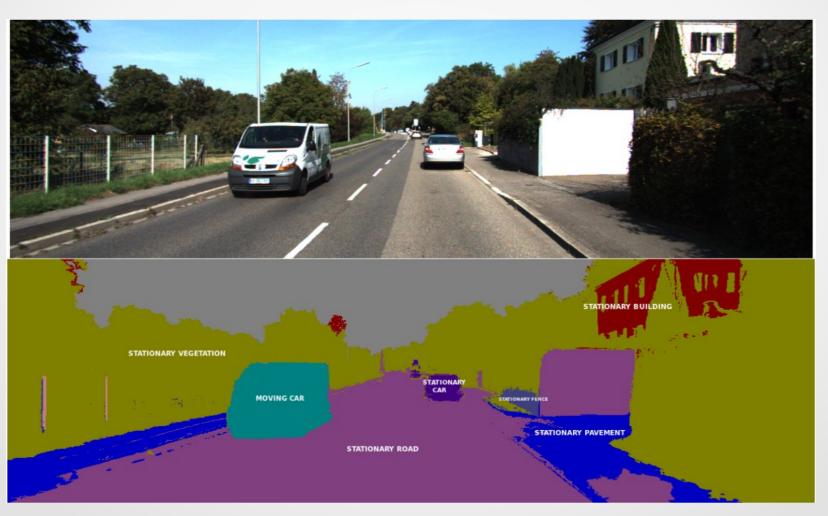
слой подвыборки (subsumpling)

слой MLP



Модель фона с помощью нейросети

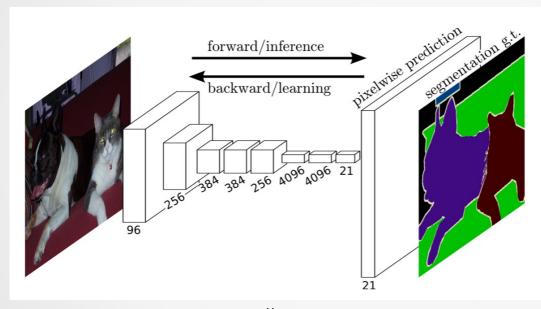
семантическая сегментация



Semantic segmentation

FCN: Fully Convolutional Networks

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





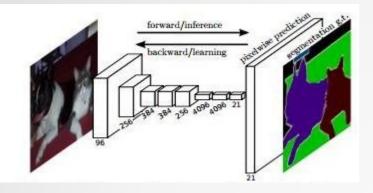


выход — карты поточечной оценки

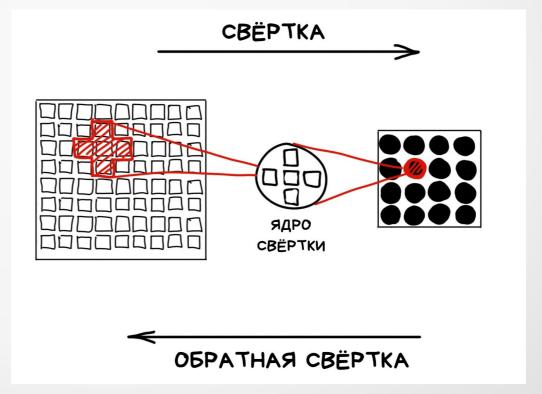
для каждого класса своя карта

размер входного изображения = размеру входной карты

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



изображение обрабатывается свёрточными слоями на выходе выполняем обратную свёртку



https://vesnins.ru/vychislitelnaya-fotografiya-budushhee-fotografii-eto-kod

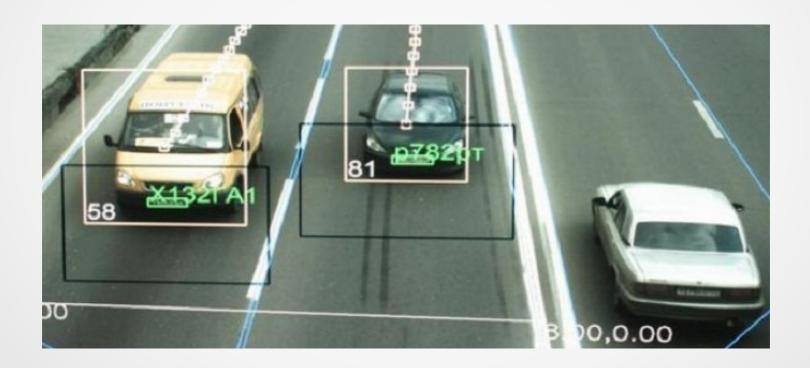
FCN: Fully Convolutional Networks

Пример — ищем людей на картинке (датасет Pascal VOC)

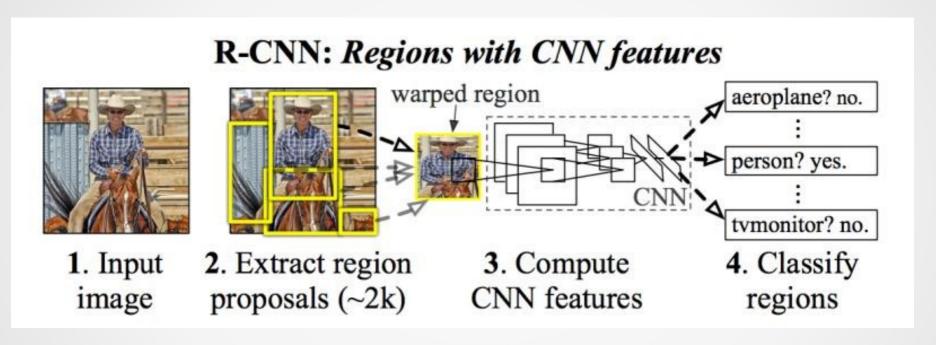


Модель объекта с помощью нейросети

Задача классификации



Object detection



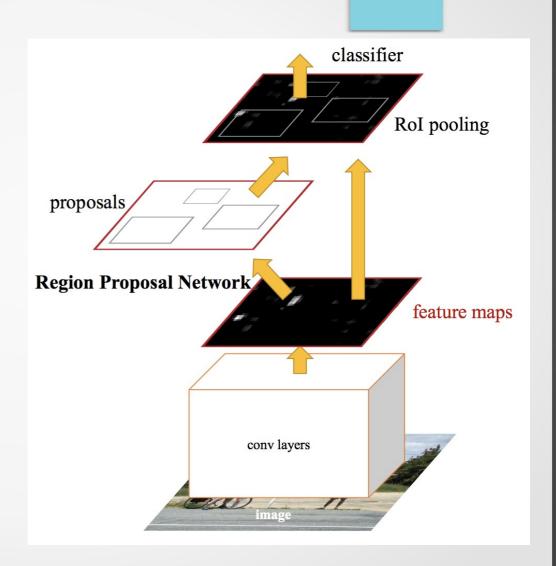
Region Based Convolutional Neural Networks (R-CNN)

изображение разделяется на части

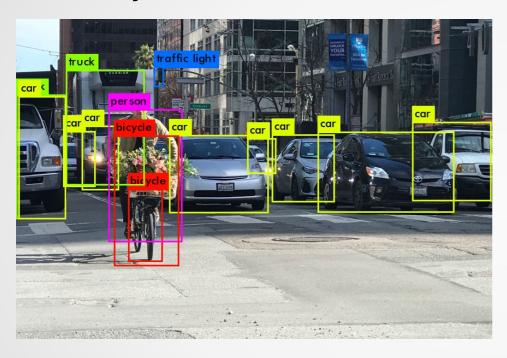
каждую часть проверяем классификатором

Faster-R-CNN

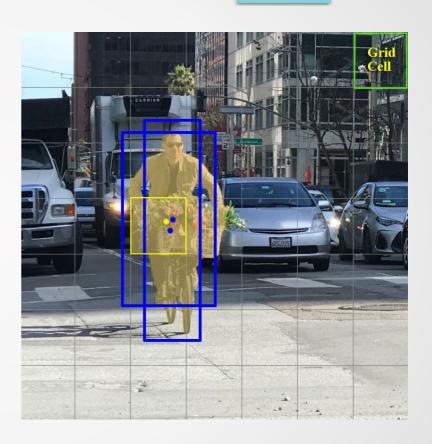
- принимаем картинку на вход
- картинка прогоняется через CNN, формируем feature maps
- определяем регионы-кандидаты (возможно содержащие объекты)
- выделяем эти регионы
- и применяем к ним классификатор картинок



Object detection



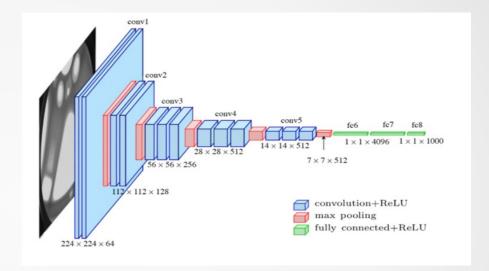
YOLO: You Only Look Once

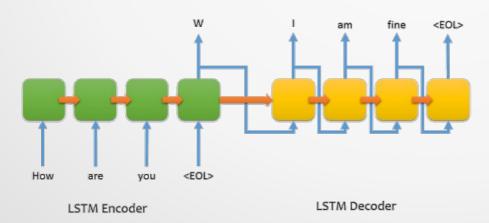


Генератор подписей к картинкам

модификация языковой модели Seq2Seq

Encoder — свёрточная сеть VGG Decoder — рекуррентная сеть GRU

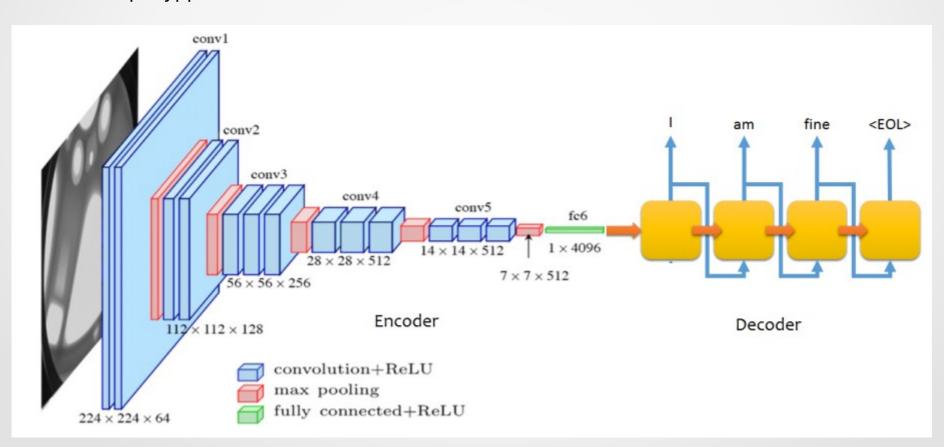




Генератор подписей к картинкам

модификация языковой модели Seq2Seq

Encoder — свёрточная сеть VGG Decoder — рекуррентная сеть GRU

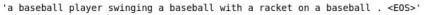


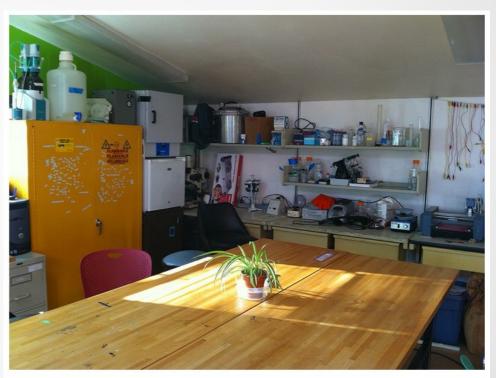
Генератор подписей к картинкам

модификация языковой модели Seq2Seq

Encoder — свёрточная сеть VGG Decoder — рекуррентная сеть GRU







'a living room with a coffee table . <EOS>'

Литература

Борисов E.C. Методы машинного обучения. 2024 https://github.com/mechanoid5/ml_lectorium_2024_I

Борисов E.C. Базовые методы обработки изображений. http://mechanoid.su/cv-base.html

Борисов E.C. О задаче поиска объекта на изображении. http://mechanoid.su/cv-image-detector.html

Конушин A.C. Введение в компьютерное зрение. 2015 https://www.youtube.com/playlist?list=PL-_cKNuVAYAXAnpy8RCV8UtFrFFLRa4rh

Николенко С., Кадурин 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