cs231n-4: Detection and Segementation

1. Segmentation （对pixel做分类）
    1. label each pixel with a category label
    2. approaches
        1. slide windows
        2. Fully convolutional without FC-layers, instead, directly give the scores of each pixel
        3. down sampling and up sampling
            1. nearest neighbor unspooling
            2. bed of nails
            3. max unpooling
                1. using the positions from pooling layer
                2. help us preserve some spatial information
            4. transpose convolution
                1. above are fixed function
                2. 用一个标量乘上一个weight matrix，和conv layer相反，重叠的部分可以相加
2. Classification + Localization （找到一个Object，然后分类、找到边界）
    1. Two loss, assume our images are annotated by both classification labels and bounding box coordinates
        1. classification loss like softmax
        2. localization loss like L2 loss
    2. Aside: Human Pose Estimation （姿势预测）
        1. outputs are 14 numbers giving the x and y coordinates
        2. regression loss is ok
        3. big point is we should know the number of outputs
    3. A complementary about loss
        1. discrete output: softmax, cross entropy or other things
        2. continuous output: L2 regression loss
3. Detection (detect boundaries) （多个Object，分类，找到边界）
    1. problem is we don't how many objects are in the image
    2. sliding window **bad**
        1. how do you choose? brute force is a bad approach
        2. computationally intracable
    3. region proposal, came up by **R-CNN, all of them are region based-methods**
        1. Find blobby regions that are likely to contain objects
            1. the regions will be a little bit higher than the origin
        2. advantages
            1. pretty good at recall
            2. really fast to run
        3. approach
            1. first apply region proposal networks to get some regions, then apply CNN to classify on each region
            2. notice that the regions' sizes are not the same, so we should wrap them to a fixed square size
            3. a classification to find out the region's category (like SVM) and a

regression to find out the boundary box of the object (like L2 loss) are used

4. disadvantages
    1. Training is still computationally expensive and slow
    2. take a lot of disk space
    3. Testing is also slow, because we should apply CNN on ~2000 regions
5. **Fast R-CNN**
    1. 我们不再用原图上的ROI (region of interest)，而是Conv layer上的ROI
    2. 瓶颈变成了region proposals
6. Faster R-CNN
    1. four loss
    2. In Region Proposal Network
    3. classification loss (binary, it's a region or not)
    4. regression loss (region boundary)
    5. In final CNN network
    6. classification loss (which class it is)
    7. the object's boundary

4. Detection without Proposals
    1. **YOLO (You Only Look Once)/ SSD (Single Shot Detection)**
        1. divide image into N * N grid
        2. use B base bounding box
            1. dx, dy, dw, dh, confidence
        3. Meanwhile, do classification to C classes on each grid
        4. total output
            1. N*N*(5*B + C)
        5. It's fast but R-CNN is more accurate

4. Instance Segmentation（多个Object，分类、找到这个Object的区域）
    1. Mask R-CNN
        1. …这个东西好无聊啊，就是在找到多个Object的边界以后，在这个小图里面做segmentation