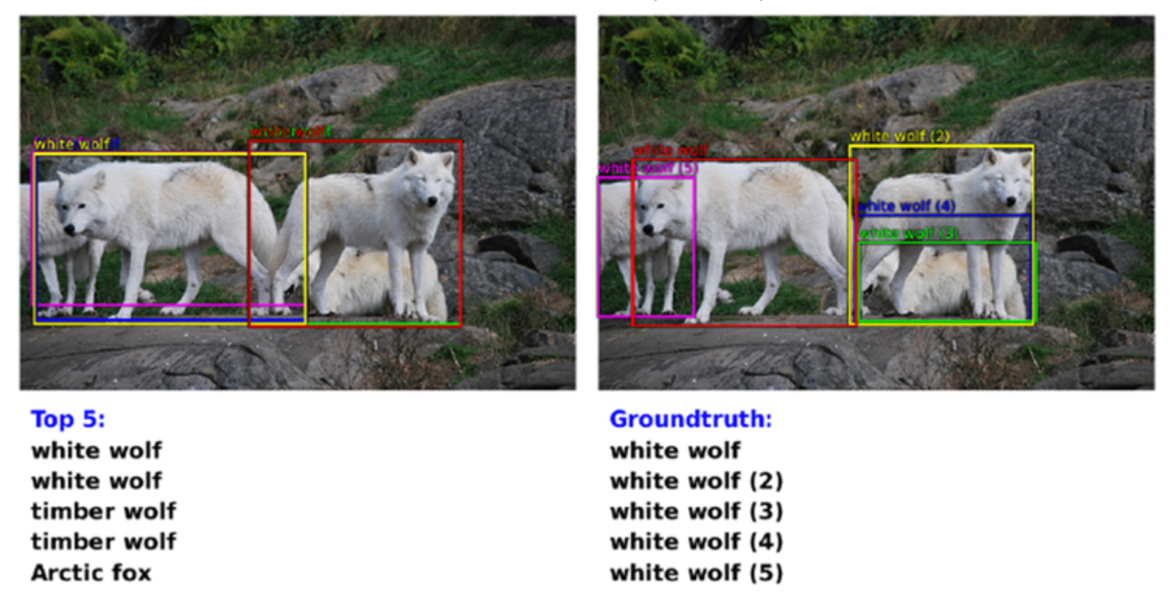
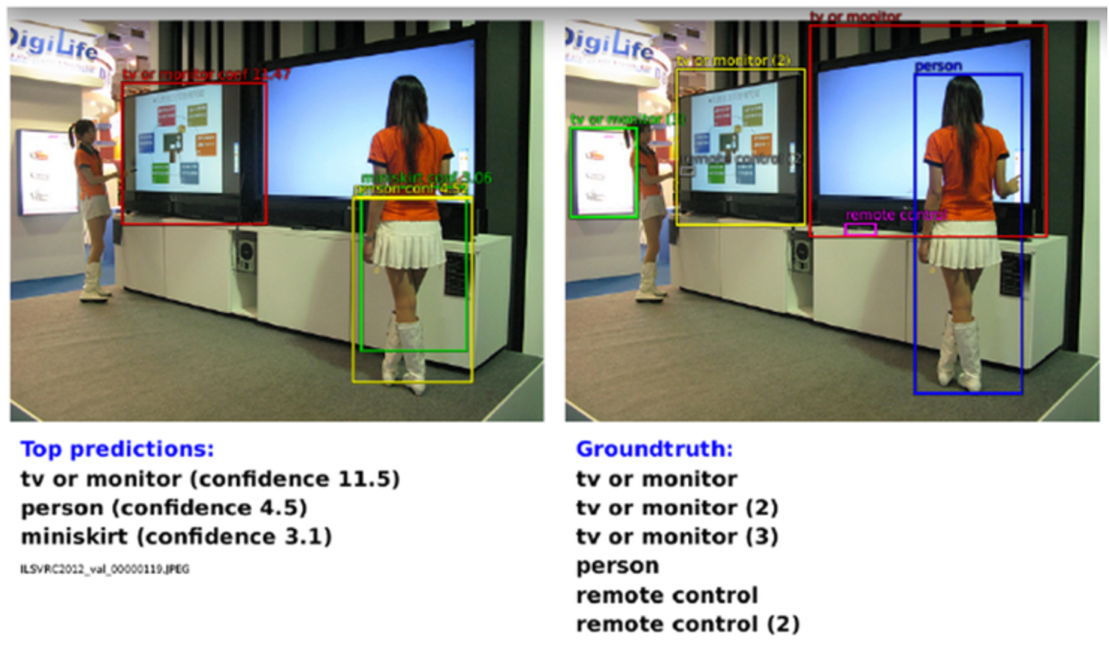
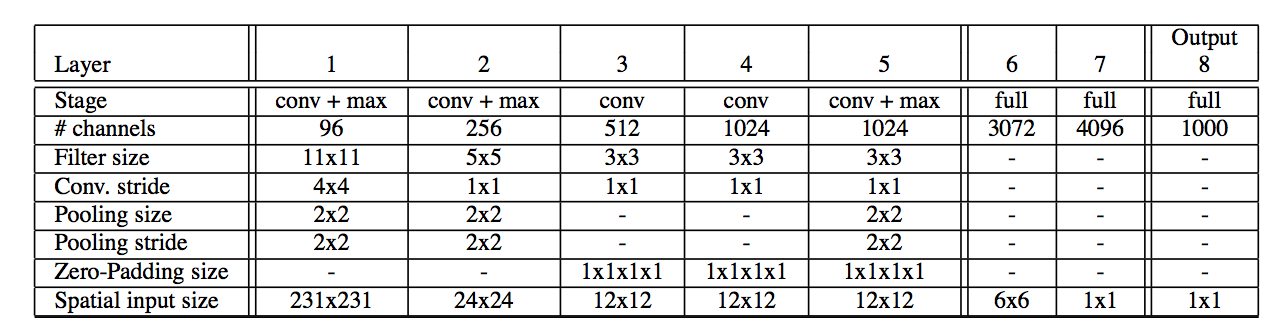
Overfeat: Integrated Recognition, Localization and Detection

Jo MinKi

1. Introduction   
   Overfeat is integrated framework for classification (determine the class of image like bear, cat..), localization (find the location of bear in the image) and detection (detect and determine the location of the object inside the image). It was big news in this field at 2013, but now there are better model for each of task. For example in classification, following researches such like Google net, VGG, ResNet get much better and in detection, R-CNN which is using selective search showed much better performance. So now Overfeat is not most popular model anymore.
2. Vision Tasks  
   They described three CV tasks in increasing order of difficulty, classification, localization, and detection.   
   Classification is that finding the existence of the specific object. Usually they represent top-5 similarity and select the top entry. Following example is that finding the Leopard in the image.   
   Localization is drawing a bounding box around the object so the location of the image can be specified. This process use the same dataset that used for classification. This process can specify up to 5 locations of objects.   
     
   Detection use 2000 class of data and infer the existence and location of the object from the image as possible as they can. So if the image have no object then it should print 0 and give a penalty when the inference get wrong.   
     
   
3. Classification  
   Basically the architecture of Overfeat is based on Alexnet, and they improved on network design and the inference step.   
   1. Model design and training   
      Basic architecture is based on AlexNet. (Following configurations are too specific and there’s no explanation about the reason, so I will just list up them)
      1. Do random crop and reflection (data augment) and present to mini batch
      2. Initialize the weight in the network randomly with
      3. Set the learning rate to initially and decrease by a factor of 0.5 after several epochs
      4. DropOut with a rate of 0.5 on the 6th and 7th layer which is fully connected layer

Other details are pretty much same with AlexNet such like max pooling, reflection and ReLU are used. But there are followings differences.

* + 1. Overlapping max pooling did not used for this.
    2. No contrast normalization used (at the AlexNet paper, this called Local Response Normalization and there are many other alternative names of it. It seems LCM(Local Contrast Normalization) is mostly used.
    3. Instead they gave up the non-overlapping max pooling (which means bigger stride), they used bigger layer feature map at the 1st and 2nd layer. This make the model gets faster but lose some accuracy.
    4. Remaining coefficients are in the following table.  
         
       First layer capture oriented edges, patterns and blobs. Filters in the second layer have a variety of forms, some diffuse, other with strong line structures or oriented edges.
  1. Feature Extractor   
     They released a feature extractor named OverFeat. OverFeat has two models, a ‘fast’ and ‘accurate’ one and each architecture will be described later.
  2. Multiscale Classification

Question.

1. At the 3.1.5, So is overlapping max pooling not a necessary process? Or why did they passed that process?  
   -> They passed just for speed, so they lost the accuracy.
2. What does the ‘Feature extractor’ do?