

## 1、L2-normalize 的计算方式

$$L2\_Norm = \frac{x_i}{\sqrt{\sum_{k=1}^n x_k^2}} \quad // \text{L2 Normalize}$$

## 2、HyperNet

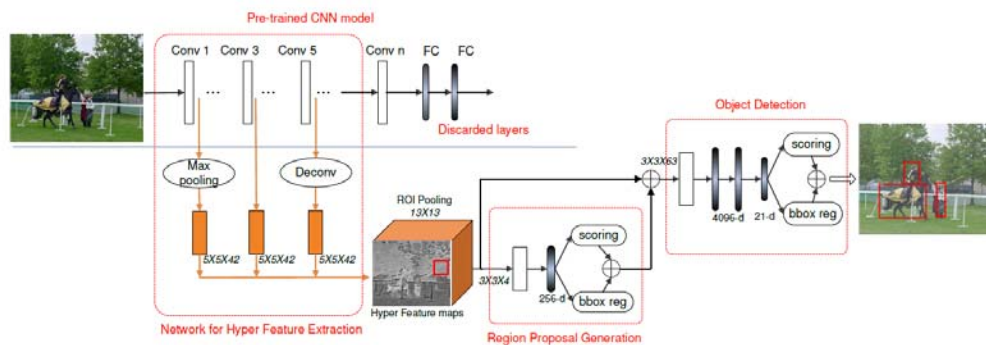


Figure 2. HyperNet object detection architecture. Our system (1) takes an input image, (2) computes Hyper Feature representation, (3) generates 100 proposals and (4) classifies and makes adjustment for each region.

这里昨天讲的有错误我把 13x13 误认为是 Hyper Feature maps 的大小。更正如下：

- (1) 论文中给出的一个例子中输入图像的大小是 1000x600；
- (2) 对应的 conv3 的大小是 250x150，最终得到的 Hyper Feature maps 的大小也是 250x150；
- (3) 建议框个数略大于 30k？

感觉论文中给出的细节不足以推断这里具体的实现方式

## 3、YOLO9000 中 boxes 数目

Boxes 的数目大于 1k，这里是怎么算出来的？

论文中对于典型的 416x416 大小的输入图像，最终得到的 feature map 大小为 13x13，采用 anchor box 的个数为 9，则一共的 boxes 的个数为 13x13x9=1521。后面论文中采用 k-means 方法进一步减少这个值。

## 4、MR-CNN

semantic segmentation 部分的 semantic segmentation aware activation maps 的分辨率是多少？

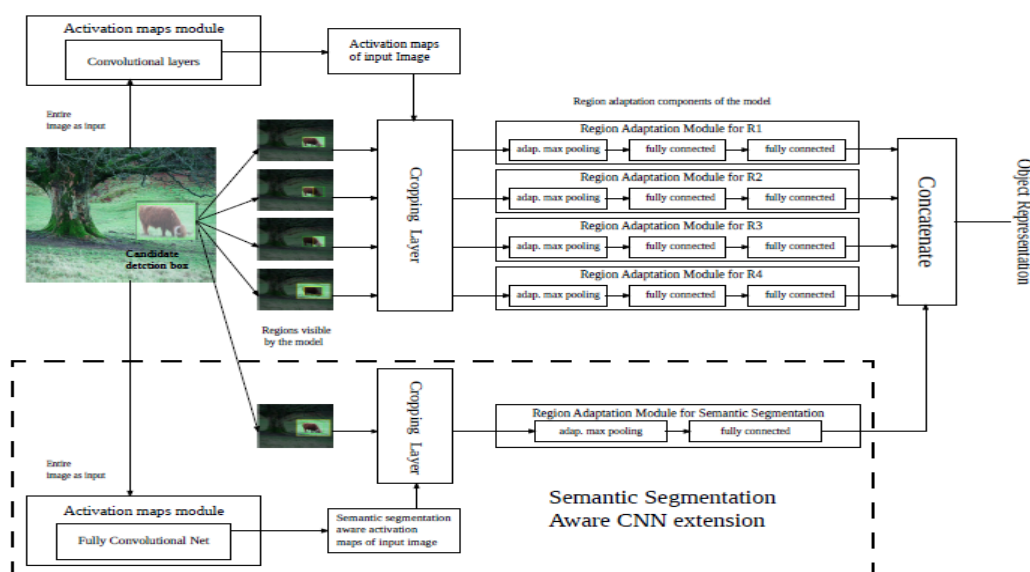


Figure 4: Multi Region CNN architecture extended with the semantic segmentation-aware CNN features.

下面图是一个语义分割网络的示意图。这里 semantic segmentation aware activation maps 指下面图中的 output 部分，其和原始输入尺寸一样。

