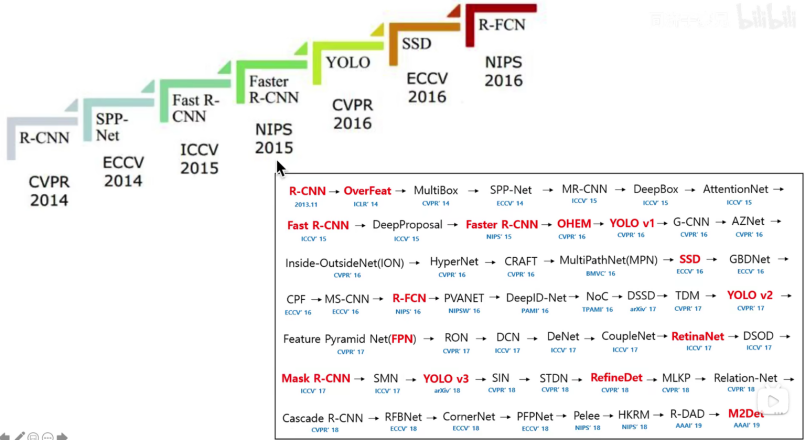
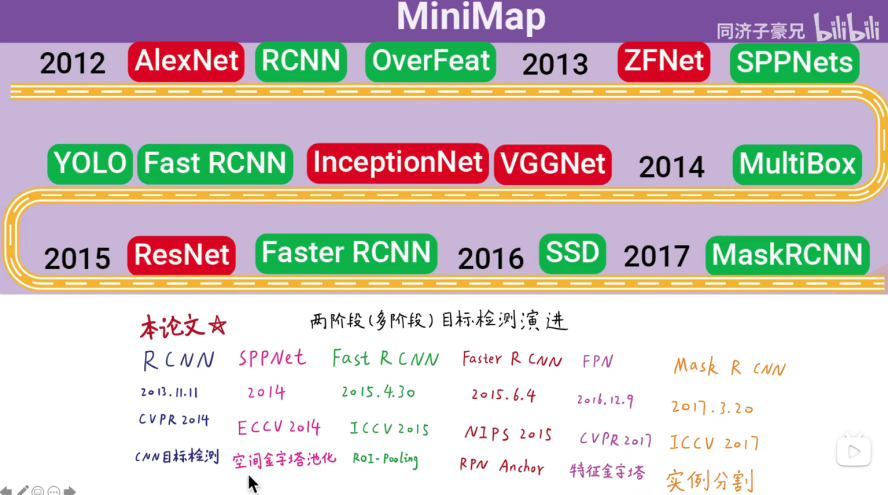
1. RCNN既可以解决目标检测问题，也可以解决语义分割问题，一般来说是解决目标检测问题



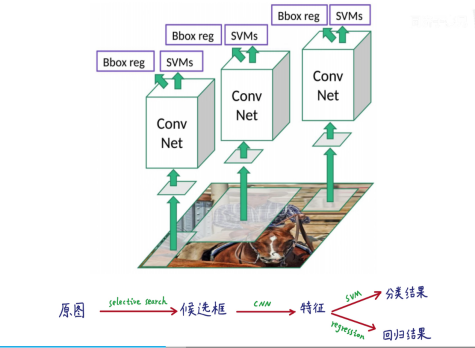
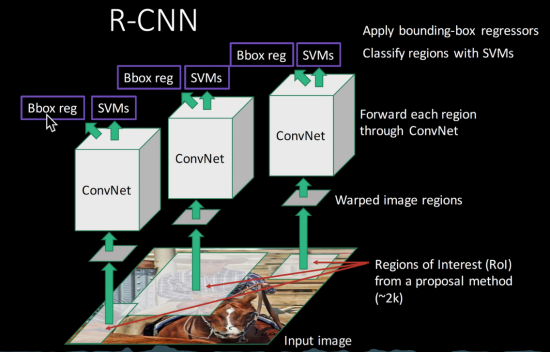
1. 红色的是分类模型，绿色的是目标检测模型



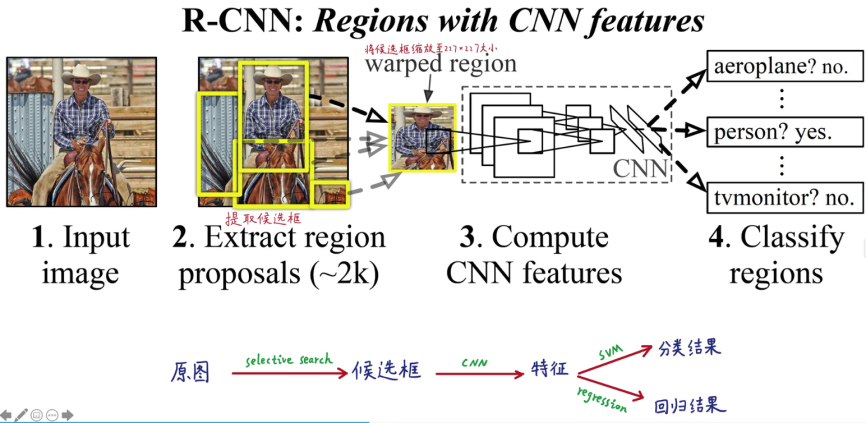
1. **工作原理**
2. .输入图片
3. .用selective search的方法在图像中生成2000个候选框，这些候选框可能是最终的目标也可能不是
4. 把每个候选框不管长宽比例大小，统一缩放成227×227大小的图片，逐一喂到同一个卷积神经网络里面，提取一个4096维全连接层输出的特征，获得这个特征之后用线性支撑向量机来对他进行分类，比如PascalVOC有20个类别，就由20个线性支撑向量机来对这个4096维的向量进行分类。这个4096维的向量，既用于线性支撑向量机的分类也可以用于bounding，box的回归。即一边进行分类一边进行回归，就完成了目标检测的任务。

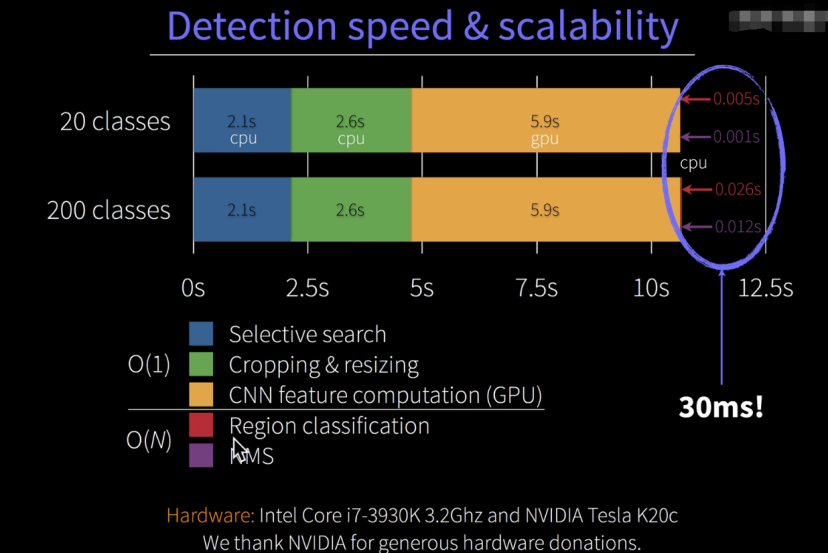
将2000×4096维特征与20个SVM组成的权值矩阵4096×20相乘，获得2000×20维矩阵，表示每个建议框是某个目标类别的得分，分别对上述的2000×20维矩阵中每一列，即每一类，进行非极大值抑制提出重叠候选框，得到该列即该类中得分最高的一些候选框。

1. 输入图像，得到每个候选框的类别，以及每个候选框精修精调过后的预测框
2. 原理图如下



1. 由上述的步骤可以看出，一环扣一环，哪一个都不能出问题，所以但凡这里面有一个步骤出问题，整个性能就毁了。每个步骤都需要单独的去优化，所以他并不是一个端到端的系统，他是一个严重依赖上下游产业链的生态





1. 由以上图片可以看出，RCNN是一个非常臃肿非常慢有严重依赖关系，一点也不端到端的网络结构。所以我们可以对RCNN进行一系列优化



**提取候选框**：比如对于提取候选框这块，我们可以使用更快的EdgeBoxes，甚至到后来的FasterRCNN中直接用RPN网络

**共享卷积运算**：之前RCNN是要把每一个候选框逐一地喂进去，这要进行2000次前向运算，在改进版本SPPNet和FastRCNN中就直接将原图喂进卷积神经网络，进行一次处理，所有的候选框共享这一次处理地FeatureMap，所以这样可以大大提高他的速度。

1. .因为RCNN中有一个全连接层，他的输出是固定的，所以才要把所有的候选框进行强制缩放，都缩放成227\*227的，但其实这是不科学的，物体会因为这个而进行形变。那么可不可以对这方面进行优化呢？

**兼容任意尺寸图象**：所以在SPPNet中用到了空间金字塔池化，FasterRCNN中用到了ROIpooling，这样就可以兼容任意尺寸的输入图片

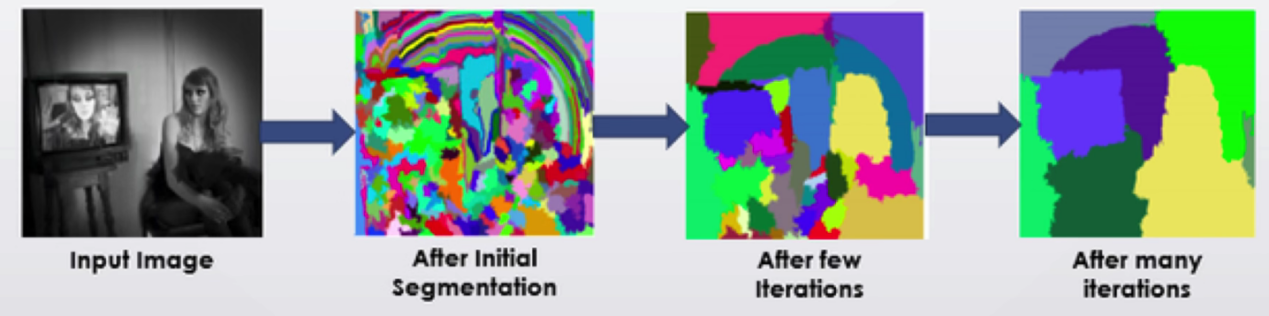
**预设长宽比**：在FasterRCNN中加了一些Anchor可以预设长宽比固定的框，在这些长宽比固定的框框上来进行检测，比如矮胖的和高瘦的长宽比就是反着的，矮胖的用来检测汽车和自行车，瘦高的用来检测行人之类的

**网络结构**：端到端

**融合各层特征**：FPN

**模块一：产生候选框Selective Search**

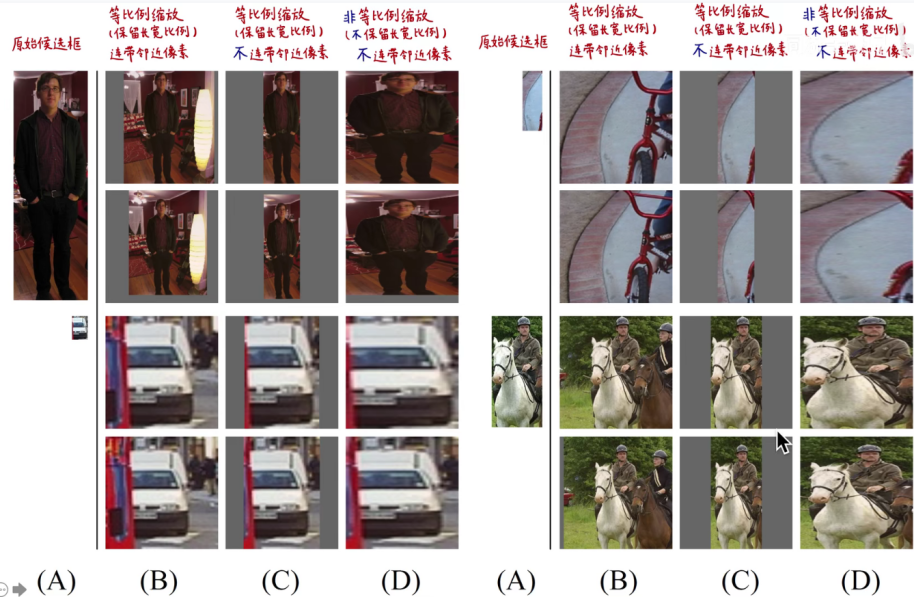
这其实是一种提取候选框的方法，他首先通过类似聚类的方法，在图像中找到一些初始的分割区域，找到一些颜色纹理大小相似度比较一致的比较相似的这些区域，然后对这些区域进行加权合并得到不同层次的2000个候选框





上图绿色的框，表示真实的GroundTruth，那么这2000个里面总有那么一两个能跟真实的物体，真实的GroundTruth对应的，所以他的Recall是比较高的

**模块二：将候选框缩放至227\*227固定大小**



最终作者选用了连带临近像素的非等比例缩放，连带p = 16个像素

对于多出来的像素，全部填充为这张图片像素的平均值，在喂到卷积神经网络之前，整张图片都减去图片像素的平均值，背景就会被减为0.

**Dilate proposal**：往外边扩一圈像素，以扩了的结果作为候选框，扩多少呢？扩16个像素

扩这个操作就是为了能够捕捉到周围上下文的信息

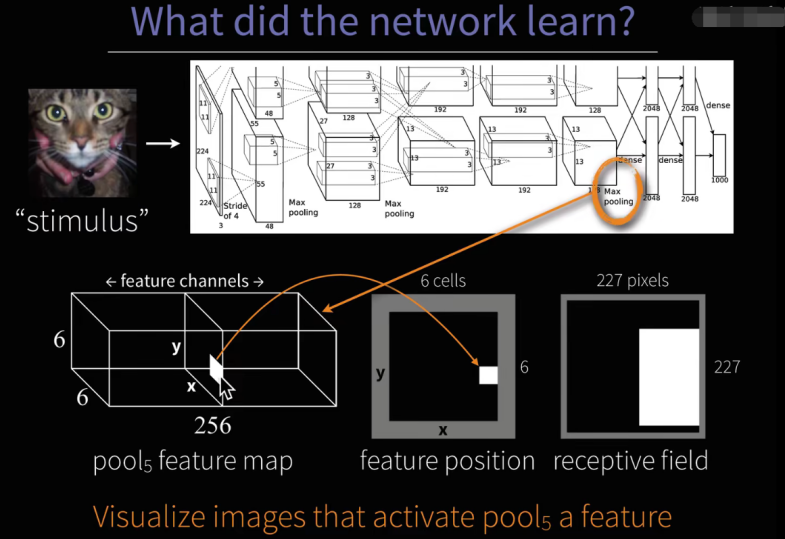


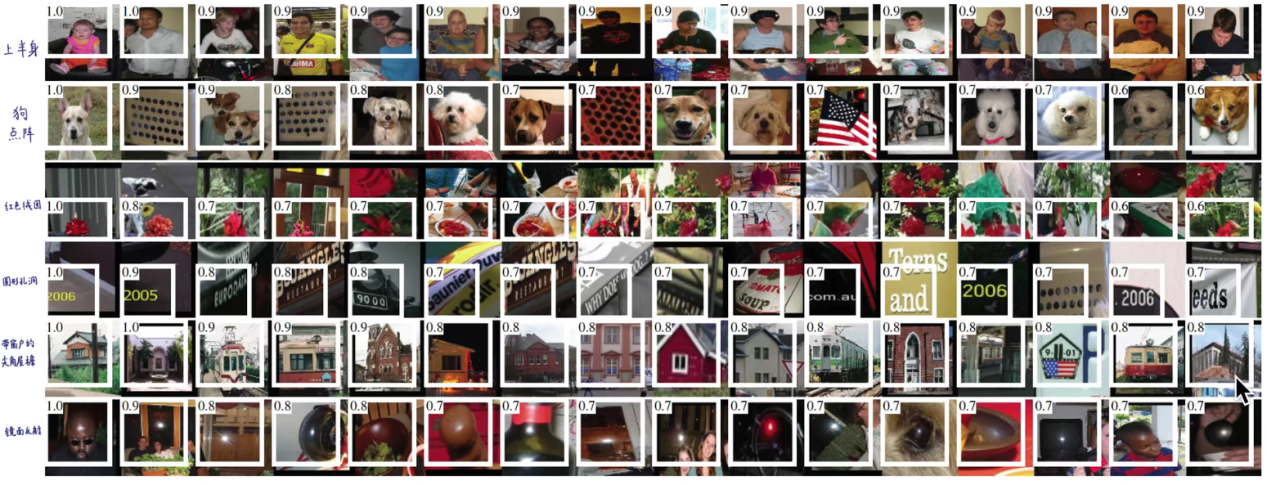
**模块三：可视化能够使得某个feature map的某个值最大化的原始候选框**

这是个什么概念呢？

我们找到某一个Chanel的某一个最大的值

这样我们就可以知道这个卷积核对原图中的哪一块区域最感兴趣



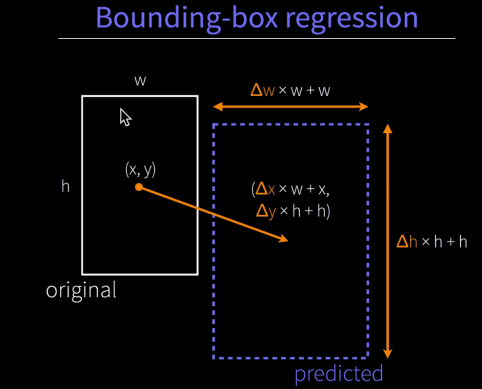


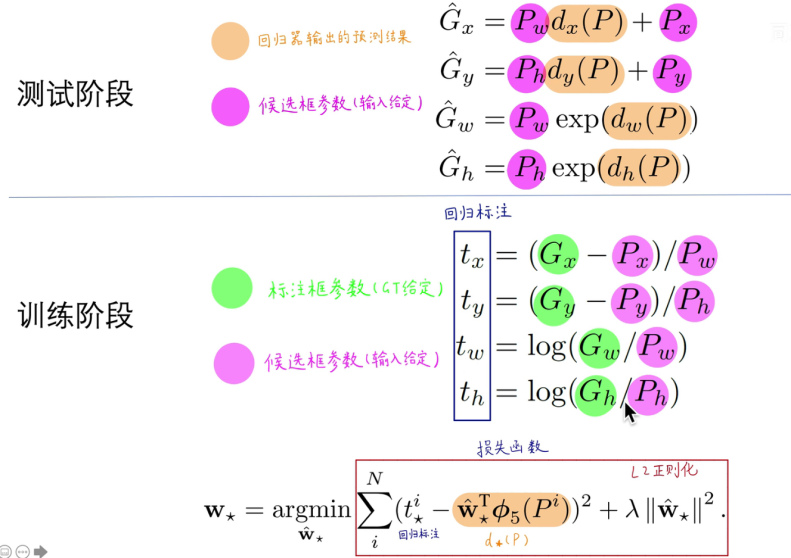
上图是从6×6×256的FeatureMap上选取了60多张Chanel来可可视化了能够使得这60多个Channel中的某一个值最大化的24个原始的region proposal（候选框）

通过这种方法，能够探究出每一个神经元提取的是什么样的特征，什么样的模式，这也是一种神经网络可解释性的方法。

**模块四：Bounding BoxRegression**

对候选框进行精调，得到一定的偏移量，然后对候选框施加这个偏移量，得到最终的预测框





**岭回归问题**

