RCNN算法流程可以分为四个步骤

1. 一张图生成1000~2000个候选区域（用selective search方法）
2. 对每个候选区域，使用深度网络（图片分类网络）提取特征
3. 特征送入每一类的SVM分类器判别是否属于该类
4. 使用回归器精细修正候选框位置
5. **生成：**

用SS算法通过图像分割的方法得到一些原始区域然后把这些区域合并，得到一个层次化的区域结构，而这些结构就包含着可能需要的物体

1. **对每个候选区域使用深度网络提取特征**

将2000候选区域缩放到227x277pixel（之所以要缩放为统一尺寸，是因为它处理的图片的尺寸必须一致），接着把候选区域输入给训练好的AlexNet CNN（去掉全连接层？）网络获取4096维的特征得到2000x4096维矩阵。

1. **特征送入每一类的SVM分类器，判定类别**

将2000x4096维特征与20个SVM组成的权值矩阵（4096x20，第一列第二列，每一列代表着要进行检测的一种物品）相乘获得概率矩阵，每一行代表一个建议框归于每个目标类别的概率，分别对上述2000x20维矩阵中每一列即每一类进行**非极大值抑制剔除重叠建议框**，得到该列技改类中得分最高的一些建议框。概率矩阵第一列代表着所有候选框为第一种物品的概率，以此类推。

1. **使用回归器精细修正候选框位置**

对NMS处理后剩余的建议框进一步筛选。接着分别用20个回归器对上述20个类别中剩余的建议框进行回归操作，最终得到每个类别的修正后的得分最高的bounding box。

处理前的是建议框（Region Proposal）理想的是实际框（Ground Truth）,回归后的窗口叫Region Proposal。

RCNN有三个模块：我们的物体检测系统由三个模块组成。第一个生成**与类别无关的区域提案**。这些建议定义了我们的探测器可用的候选检测集。第二个模块是一个大型卷积神经网络，它从每个区域抽取一个固定长度的特征矢量（4096维）。第三个模块是一组特定于类的线性SVM。

CNN：high-capacity convolutional neural Net works（大容量卷积神经网络），之前用于图像分类。

配合上区域建议（Region Proposals）结合为RCNN可用于图像目标检测

RCNN的训练策略：采用监督预训练+领域微调策略：先在大型分类数据集（large dataset ILSVRC）上预训练CNN（supervised pre-training），再在目标检测数据集（small dataset PASCAL VOC）上微调。

IoU（Intersection over Union）= IMG_256

非极大值抑制剔除重叠建议框:寻找最高得分的框，计算其他目标和该最高分目标的IoU值，删除IoU大于阈值的目标

问题：

1. 到底什么是region proposal（候选区域，被SS算法生成2000份）
2. 大型分类数据集ImageNet,目标检测数据集PASCAL VOC
3. 什么是选择性搜索（Selective Search）：**Selective Search，**是一种候选区域生成算法，用于生成**类别无关的区域提议（Category-Independent Region Proposals）**的算法，由J.R.R. Uijlings等人在2012年提出。它的核心目标是为目标检测任务高效地提供可能包含物体的候选区域（Region Proposals），从而避免传统的滑动窗口（Sliding Window）方法计算量过大的问题。然而它也存在问题，那就是过于依赖图像分割，所以速度较慢。
4. 什么是AlexNet（AlexNet是卷积神经网络网络的一种，在RCNN中用来提取各个候选区域的特征，形成一个矩阵，它有五个卷积层，三个全连接层，中间还有一些池化层，原始的AlexNet是基于imageNet进行训练的，因此它在最后一层会输出一个imageNet的1000个图像分类类别，但我们只需要它提取图像特征，所以我们并不需要第三个全连接层。）
5. 什么是仿射变换（Affine Warping）
6. IoU更加具体的解析
7. 什么是mAP（mean average precision，平均精确度）
8. 什么是SVM