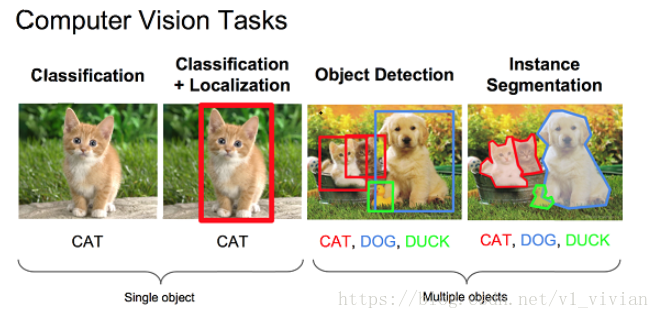
## 目标检测背景



分类问题：是什么？

定位问题：是什么，在哪里？（单一目标）

检测问题：分别是什么，各自在哪里？（多重目标）

RCNN即region with CNN features，是现在工业级应用的解决方案，随后还有Fast-RCNN ，Faster-RCNN都沿袭R-CNN的思路。

## RCNN

RCNN的方法结合了两个关键因素：

1. 对于候选区域，至下而上使用CNN定位和分割物体。
2. 通过辅助任务的预训练，再迁移到特定任务上fine-tuning，以解决训练样本不足的问题。

其关键在于将region proposal和CNNs结合起来。

#### 理论步骤

###### 区域推荐

目的是产生类别无关的区域候选，RCNN对区域推荐的算法没有要求，本文采用选择搜索法和先前的检测工作进行控制变量的对照。

###### 特征提取

作者使用CNN网络从的图片抽取出4096维特征。实际上候选的区域大小并不是固定的，作者直接采用resize的方式进行变换。

###### 分类

将由CNN提取出来的特征输入SVM进行分类。

#### NMS

non-maximum suppression（非极大值抑制）在目标检测中是为了去除冗余的候选框。

###### 原理

NMS根据重叠区域面积比例(IoU)选择最佳的区域。首先我们具有两个集合和，中为所有的候选框和对应置信度，为最终的检测框，最初为空，阈值为。

每轮将中置信度最高的候选框加入并从中移除，随后在中选出与重叠比例大于的候选框删除。反复执行直到为空。

###### 应用

NMS运作在两个部分。

在区域推荐时，推荐会产生大量重叠的区域，这些区域都对应同一个object（类别无关）。为了降低后续计算的耗费，通过NMS把最优的tos-N区域筛选出来。

在最终的检测时，NMS还会更加细分，对每一个类(per-class manner)单独应用NMS，也就是说同样一系列区域，NMS会根据per-class选取不同的区域出来。

#### 实践步骤

1. 训练（或者下载）一个分类模型（比如AlexNet）
2. 对该模型做fine-tuning

* 将分类数从1000改为21，比如20个物体类别 + 1个背景
* 去掉最后一个全连接层

1. 特征提取

* 提取图像的所有候选框（选择性搜索Selective Search）
* 对于每一个区域：修正区域大小以适合CNN的输入，做一次前向运算，将第五个池化层的输出（就是对候选框提取到的特征）存到硬盘

1. 训练一个SVM分类器（二分类）来判断这个候选框里物体的类别，每个类别对应一个SVM，判断是不是属于这个类别，是就是positive，反之nagative。
2. 使用回归器精细修正候选框位置：对于每一个类，训练一个线性回归模型去判定这个框是否框得完美。

## Fast-RCNN

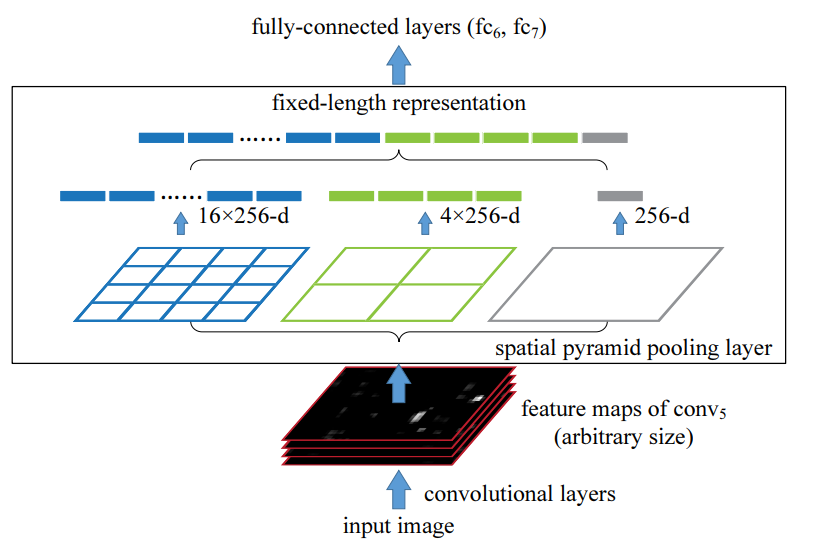
RCNN的区域推荐部分会产生2000个候选框，这个数量过于庞大。考虑到这些区域本质上都是源图像的一部分，那么考虑先对源图像做特征提取，再通过候选框找到对应的特征区域，作为后续操作的输入。但是之前候选框大小不一致的问题我们靠resize解决了，但是现在面对的是提取后的特征，就不能随意使用resize了。

#### SPP Net

Spatial Pyramid Pooling（空间金字塔池化）

总所周知，CNN分为卷积和全连接层，卷积层对图片尺寸没有要求，全连接层则需要固定的输入长度，对于变长输入就需要切割操作。SPP的作者就想到了创造一种结构，从卷积层的输出中取得固定的输出。

对于任意输入的图像，根据图像尺寸设置池化层，使得池化层输出的尺寸为，那么这一池化层可以输出长度为的向量，按照金字塔模型，设置多层输出尺寸不同的池化层，最后将输出向量拼接起来，得到最后定长输出。参考下图



#### ROI池化

ROI(region of interest)本质上是对SPP的简化，其对不同大小的推荐区域使用不同尺寸的pooling，最后下采样到固定的的特征图

#### 优化小结

相对于RCNN，Fast-RCNN做了如下优化

1. 将RCNN的结构调整，只进行一次特征提取操作，并使用ROI池化解决了区域大小不一致的问题，这样区域的特征本质上共享自最初进行的CNN特征提取。
2. 损失函数使用多任务损失函数，将边框回归的部分加入CNN中进行训练，利用soft-max替代SVM，将分类操作合并到CNN中，使得其变成一个端到端的模型。多任务损失使得两个任务可以共享卷积抽取的特征。

## Faster-RCNN

Fast-RCNN也存在一些瓶颈：搜索候选框时使用selective search，这个过程非常耗时。

作者引入了Region Proposal Network负责提取边缘信息，进而完成候选框的选择。

#### Region Proposal Network

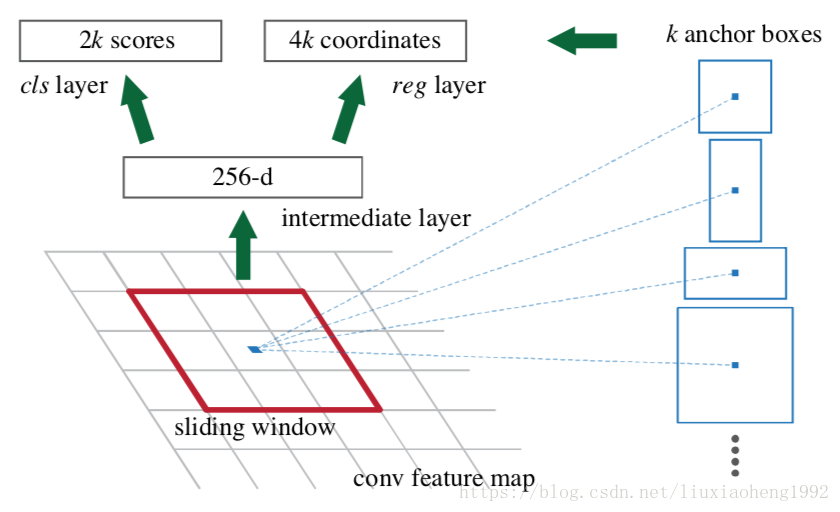
为了预测候选框，作者引入RPN来替代选择性搜索。首先遇到的问题是，如何实现大小长宽比不固定的边框估计，传统的回归估计四个端点坐标显然不合适，作者提出了anchos方法。

###### Anchos

Anchos是固定尺寸的边界框，其目的是让真正的候选框由anchos形变而来。具体一点，对于输入特征，每一个像素点都有个anchos（其包含了预设的不同比例和面积），对这些anchos进行长宽的形变，以产生候选框和相应的分数。只要获知了特征提取网络的结构，候选框可以很轻松地复原为对源图像的box。

具体而言

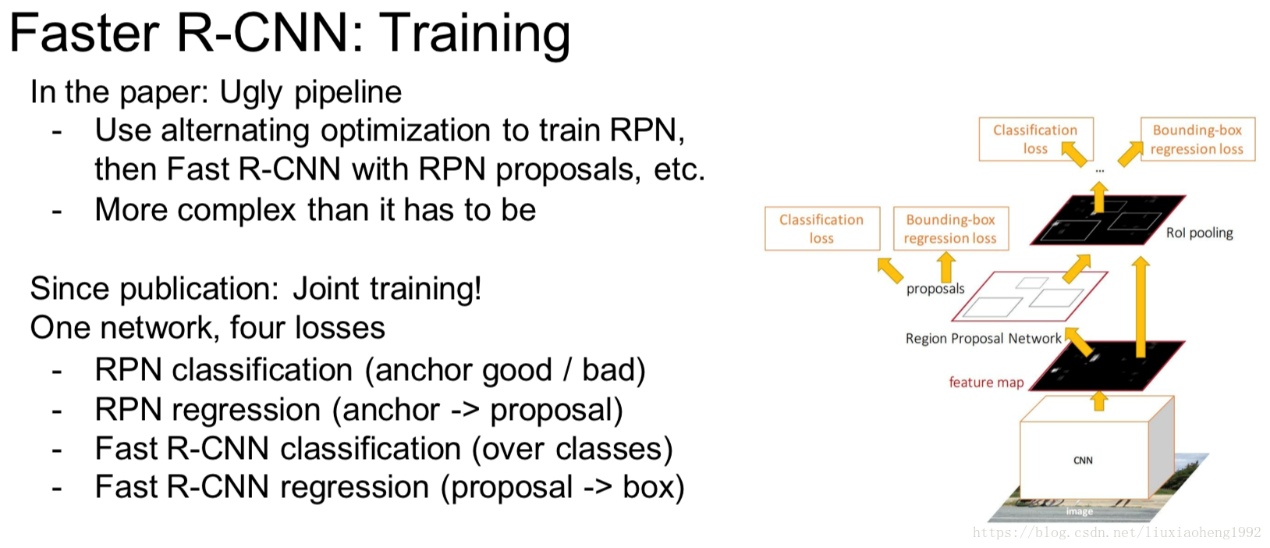
1. 对于输入特征，先做的卷积使得每个像素具有周围的特征信息。
2. 每个点具有个anchos，那么对其做像素级分类（区分前景背景）得到个分数，此外还需要对这四个偏量进行回归（理论上可以拟合任意边界框）得到个值。
3. 边界框与ground truth box地IoU大于0.7作为正样本，小于0.3作为负样本。



#### 训练

Loss函数为：

其中表示batch中第个anchor，表示前景概率，根据其真实情况标记为。表示估计坐标，表示真实坐标。



在训练和测试中，区域推荐和后续的分类（目标分类）及回归（检测框修正）。