目标检测

作者：Calvin
QQ: 179209347
Mail: 179209347@qq.com

介绍

笔记简介：

- 面向对象：深度学习初学者
- 依赖课程：**线性代数，统计概率**，优化理论，图论，离散数学，微积分，信息论

知乎专栏：

https://zhuanlan.zhihu.com/p/693738275

Github & Gitee 地址：

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

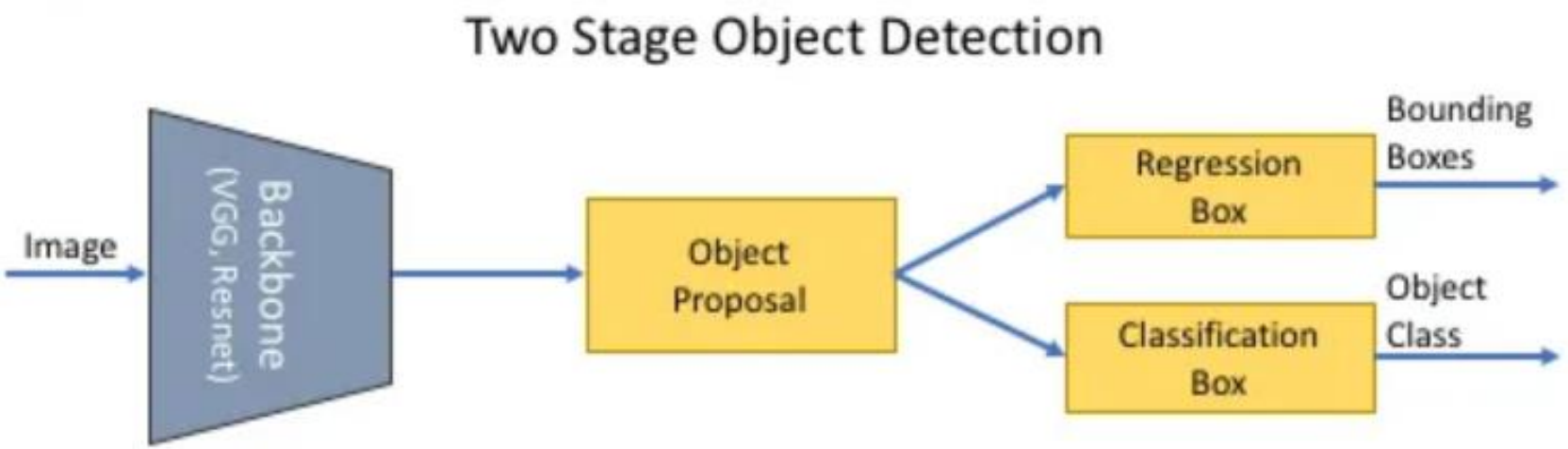
*** 版权声明：**

- 仅限用于个人学习
- 禁止用于任何商业用途

两阶段目标检测

两阶段目标检测器将整个过程分为2个步骤：

步骤1：首先使用卷积神经网络（CNN）提取特征。
步骤2：然后提取一系列感兴趣区域，称为目标候选区域，然后分类和定位仅在目标候选区域上进行。



两阶段目标检测 - R-CNN

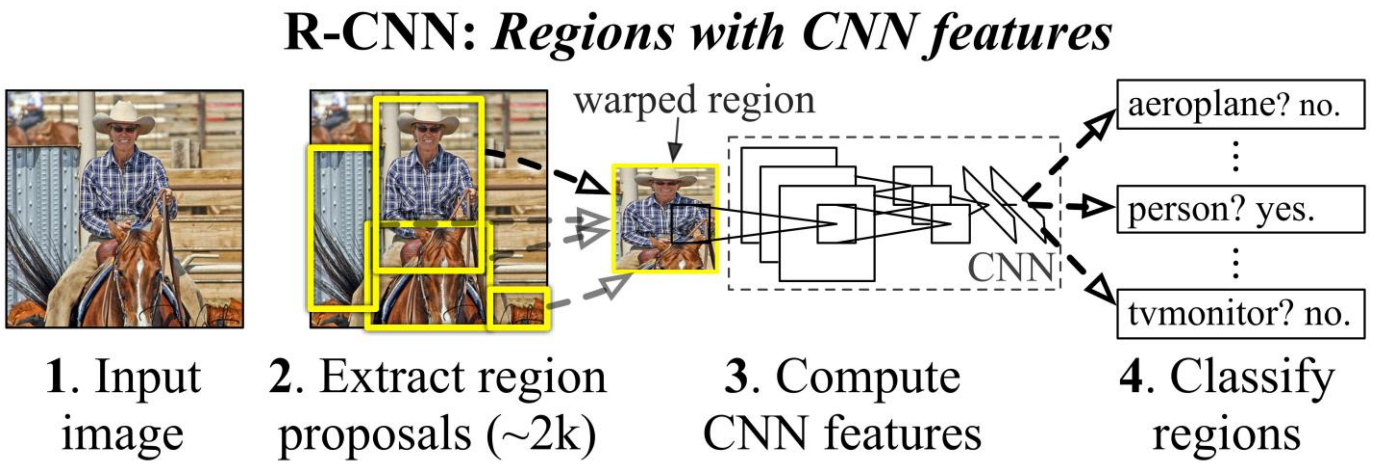
AlexNet主要用于图像分类，对于更高要求的图像识别，自然而然的想法是：从图片中获取识别对象区域，将该区域作为分类网络的输入，得到图像识别的目的。在整个过程中，主要有两个部分的工作：

- 待识别对象区域（RoI）获取
- 待识别对象区域分类

R-CNN首先通过选择性搜索算法产生候选目标，然后通过卷积神经网络对每个候选框进行特征提取，最后对候选区域特征使用SVM和回归器进行分类和回归。

R-CNN存在的问题：

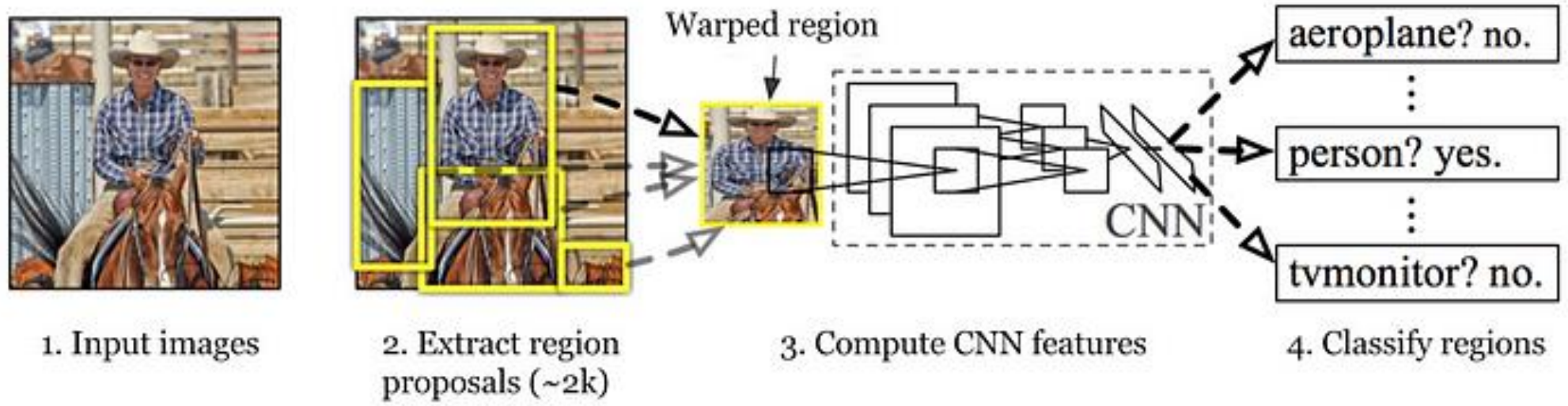
- R-CNN是需要多阶段训练才可以完成，需要很长的训练时间；
- R-CNN涉及使用全连接层，因此要求输入尺寸固定，这也造成了精度的降低；
- 候选区域需要缓存，占用空间比较大。



两阶段目标检测 - R-CNN

模型结构

- **预训练 CNN**：使用 ImageNet 等（例如 VGG 或 ResNet）预训练 CNN 模型
- **生成候选区域**：利用一些外部算法（例如选择性搜索）来建议图像中的感兴趣区域
- **Warp Region Proposals**：将每个区域提案扭曲为 CNN 所需的固定大小
- **在候选区域上微调 CNN**：在分类任务的扭曲区域建议上微调预训练的 CNN
- **特征提取**：对于每个候选区域，通过微调的 CNN 进行前向传播，生成特征向量。
- **二元 SVM 分类**：为每个类独立训练一个二元支持向量机 (SVM)
- **边界框回归**：使用 CNN 特征训练回归模型来预测边界框校正偏移。



两阶段目标检测 - SPP-net

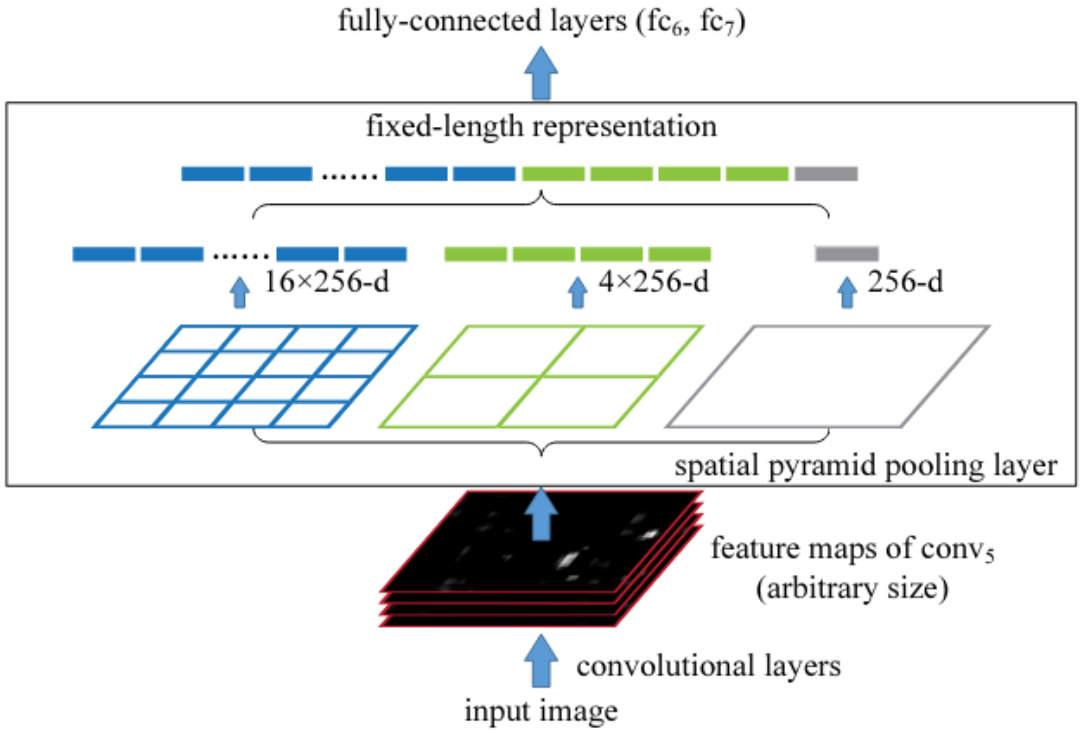
SPP-net（Spatial Pyramid Pooling Networks）是一种用于图像分类的深度学习架构，由何恺明等人于2014年提出。SPP-net的主要贡献是引入了空间金字塔池化（Spatial Pyramid Pooling）层，使得网络可以接受任意尺寸的输入，并且在不改变网络结构的情况下，能够生成固定长度的特征表示。

优点：

- 比R-CNN模型快得多，具有相当的准确性。
- 可以处理任何形状/宽高比的图像，从而避免由于输入扭曲而导致的目标变形。

缺点：

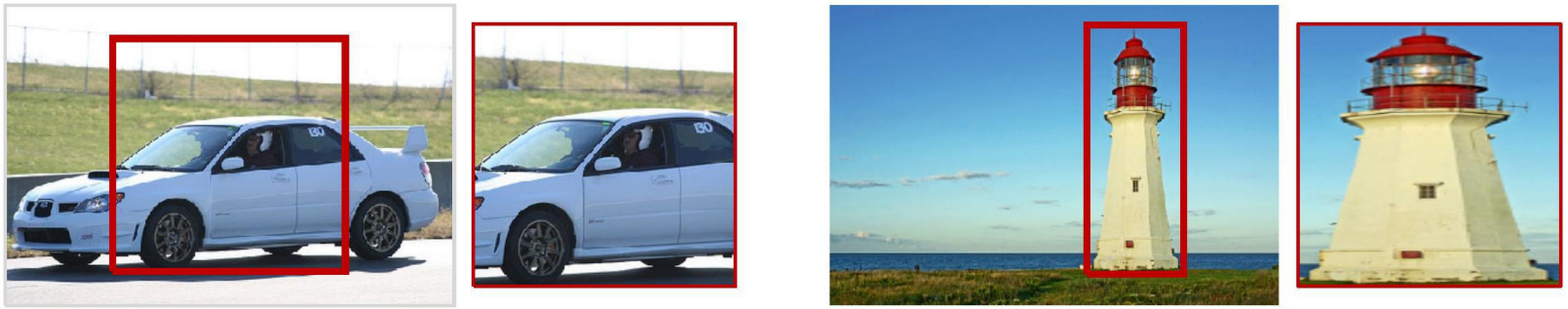
同样具有R-CNN的缺点，例如多阶段训练、计算量大和训练时间长等问题。



两阶段目标检测 - SPP-net

SPP-Net与经典CNN最主要的区别在于两点：

- 不再需要对图像进行crop/wrap这样的预处理，原始图像经过缩放之后，变得很畸形失真，这也会影响到特征提取的过程；
- 在卷积层和全连接层交接的地方添加空间金字塔池化层（spatial pyramid pooling），是SPP-Net网络的核心所在。SPP-Net在最后一个卷积层后，接入了金字塔池化层这种方式，可以让网络输入任意的图片，而且还会生成固定大小的输出。



两阶段目标检测 - SPP-net

模型结构

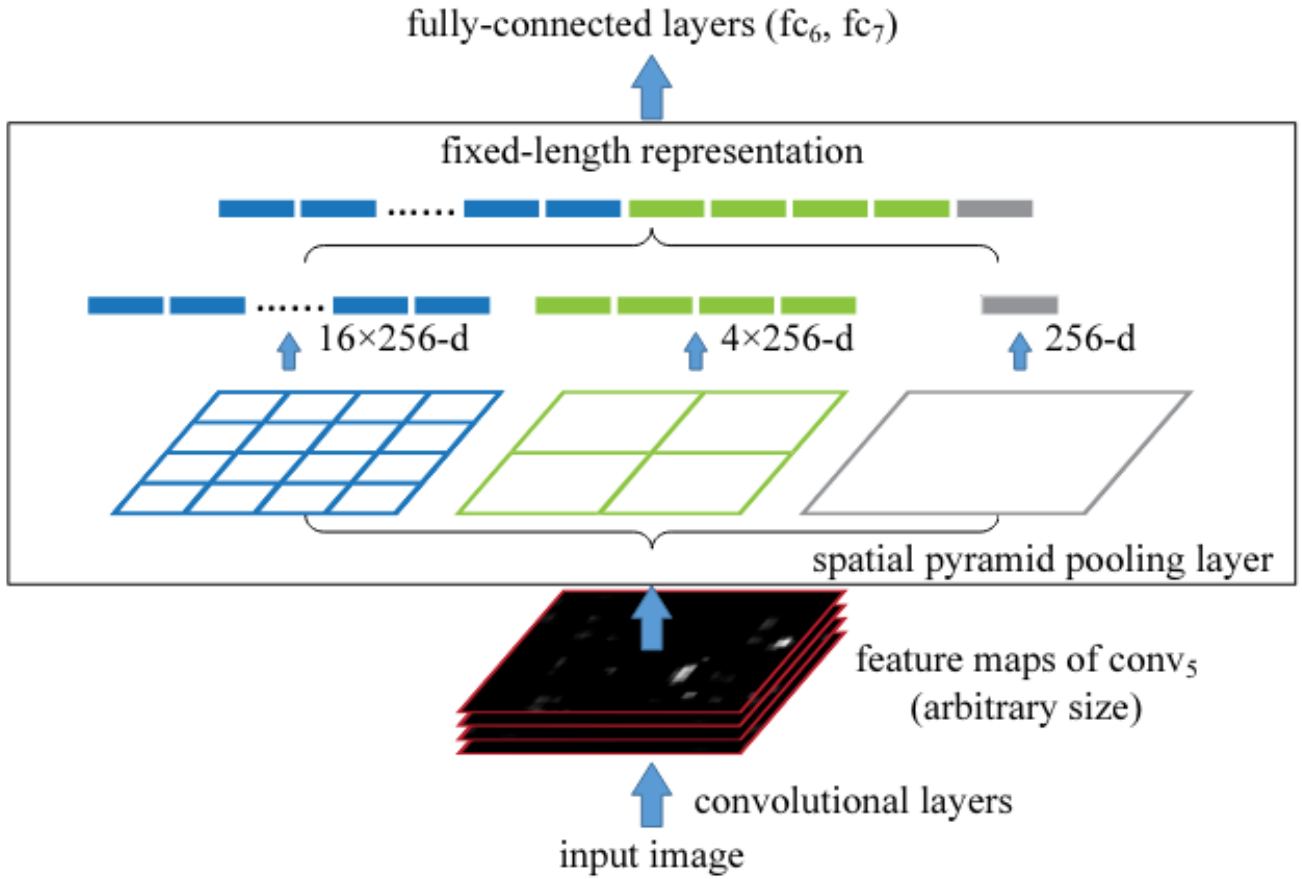
① Image经过一次卷积之后会得到256个特征图（黑色堆叠图）

② 然后使用SSP结构对这256个特征图进行处理，将这256张特征图分别进行1*1，2*2，4*4的最大池化。将三个特征向量连起来长度为 21（21=4*4+1*1+2*2）。就可以保证将原始图像中的不同长宽的区域都对应到一个固定长度的向量特征。

③ 最后我们会得到 21*256=5376 个数值作为图像高度抽象的特征

④ 之后将这5376个特征送入全连接神经网络

⑤ 最后用SVM分类器输出bounding box的x,y,w,h以及每一个bounding box的图象分类的结果

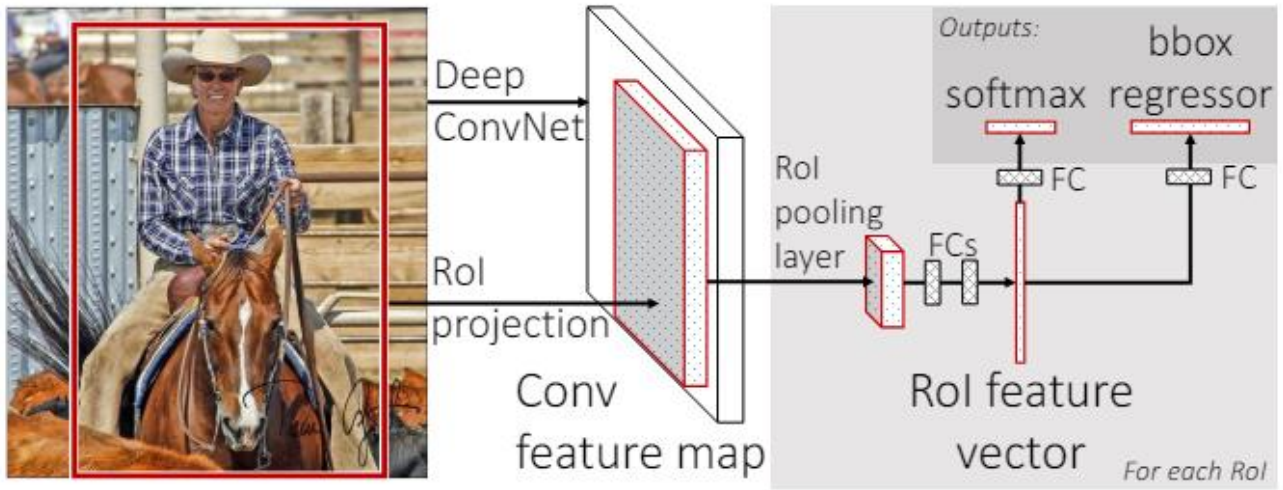


两阶段目标检测 - Fast R-CNN

R-CNN虽然完成了使用神经网络来做图像识别的开创性工作，但其存在两个主要的缺点：

- 耗时的分块处理；
- 三个模块之间分别进行训练，没有统一在一起。

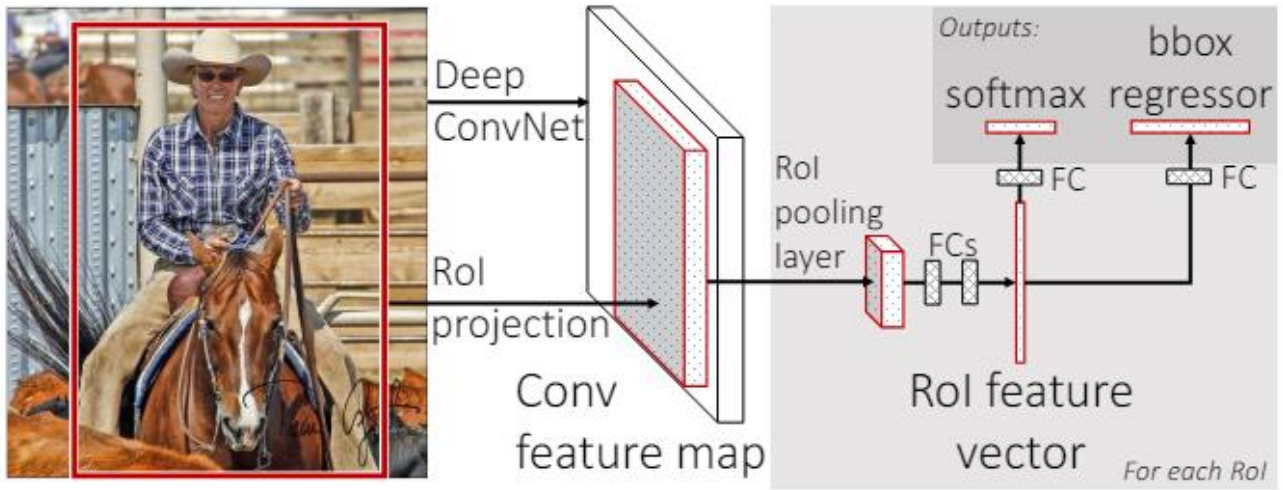
因此，Girshick 在 2015 年引入了一种名为 Fast R-CNN 的方法，以提高 R-CNN 模型的速度和效率。
Fast R-CNN 没有将目标检测视为三个独立的任务，而是将它们统一到一个框架中。
这意味着 Fast R-CNN 不是独立训练特征提取、对象分类和边界框回归模型，而是将它们组合成一个内聚系统。
Fast R-CNN的提出是为了提高速度，Fast R-CNN简化了训练过程，移除了金字塔池化并引入了新的损失函数。



两阶段目标检测 - Fast R-CNN

模型结构

- 输入图像和多个感兴趣区域 (RoI) 被输入到一个全卷积网络中提取特征；
- 对每个RoI提取的特征经过RoI Pooling层固定大小；
- RoI Pooling层出来的固定大小数据通过全连接层（FC）映射到一个特征向量；
- FC层输出的特征向量分别经过一个FC层，输出每个RoI的概率和每个类别的边界框回归偏移量。



相对于 R-CNN 改进点：

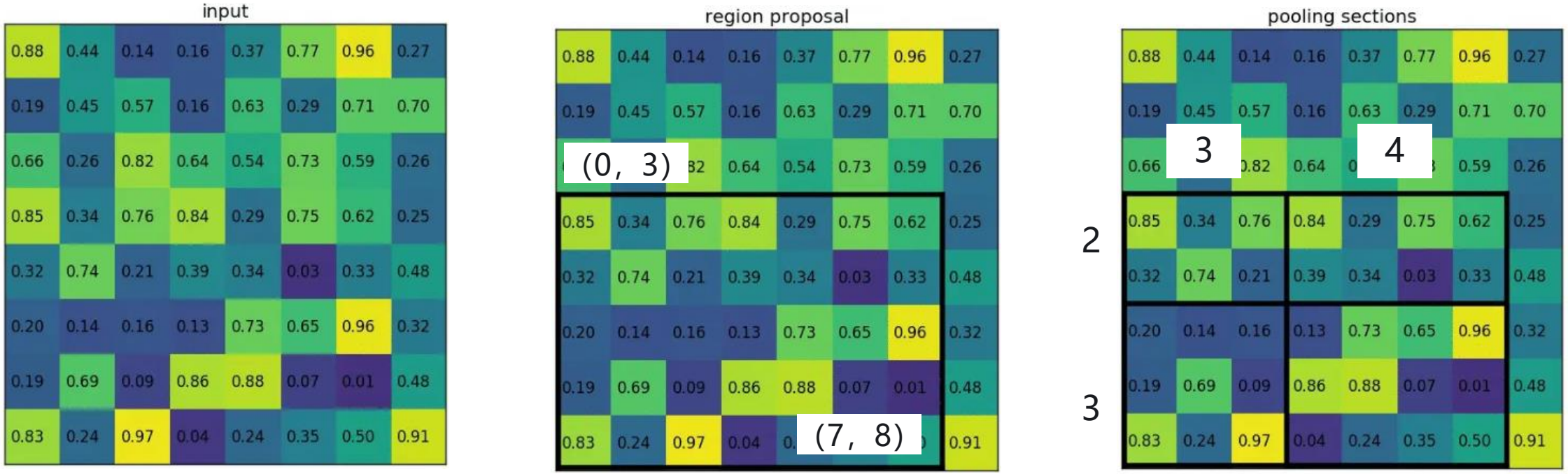
- Fast R-CNN引入了RoI池化层，可以在整个特征图上共享计算，避免了对每个候选区域进行单独的特征提取，从而提高了计算效率；
- Fast R-CNN将整个网络连接成一个端到端的结构，可以一次性对整个网络进行训练，而不需要像R-CNN那样分阶段训练；（候选区域生成仍然是独立的）

两阶段目标检测 - Fast R-CNN - ROI Pooling

感兴趣区域池化（Region of interest pooling）（也称为**RoI pooling**）是使用卷积神经网络在目标检测任务中广泛使用的操作。其目的是对非均匀尺寸的输入执行最大池化以获得固定尺寸的特征图。

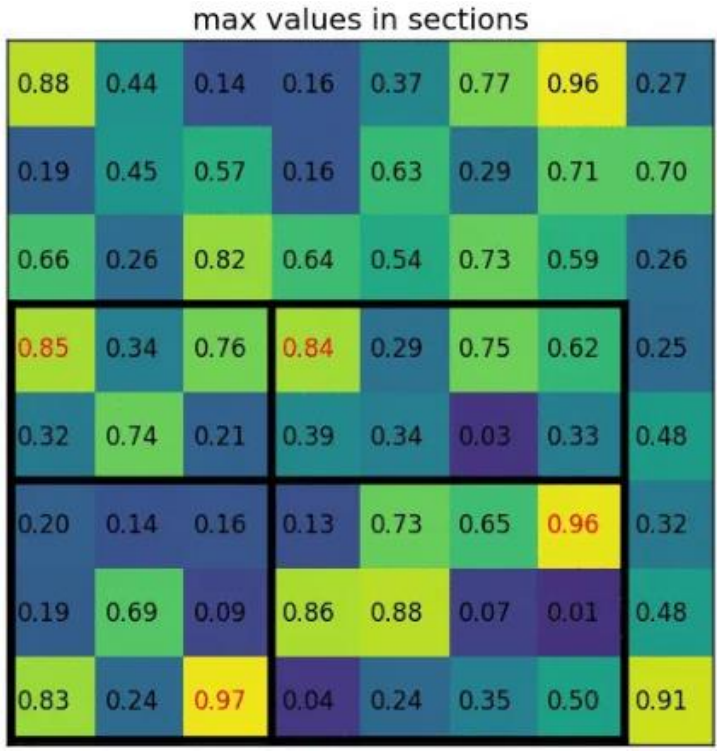
下图为一张8*8的feature map，选取其中一个5*7 的region输入ROIPooling 输出2*2结果。

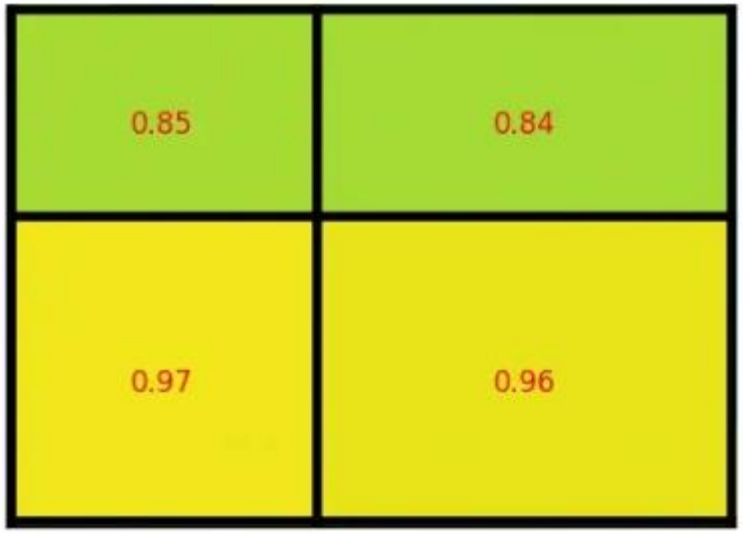
- 输入的固定大小的feature map 8*8
- region proposal 投影之后位置（左上角，右下角坐标）：（0，3），（7，8）。
- 将其划分为 2*2 块区域（因为输出大小为 2*2 ）
  1）5/2 = 2.5 -->2，剩下的为3
  2）7/2 = 3.5 -->3， 剩下的为4



两阶段目标检测 - Fast R-CNN - **ROI Pooling**

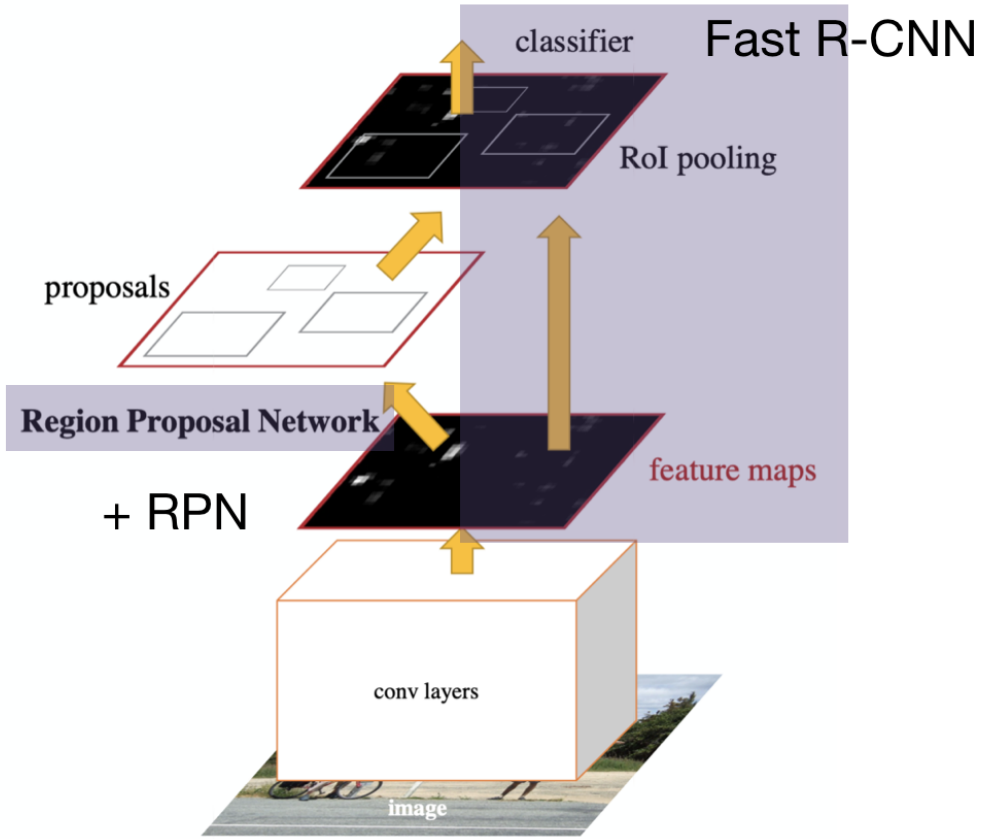
- 找到每个部分的最大值
- 将这些最大值复制到输出(max pooling)

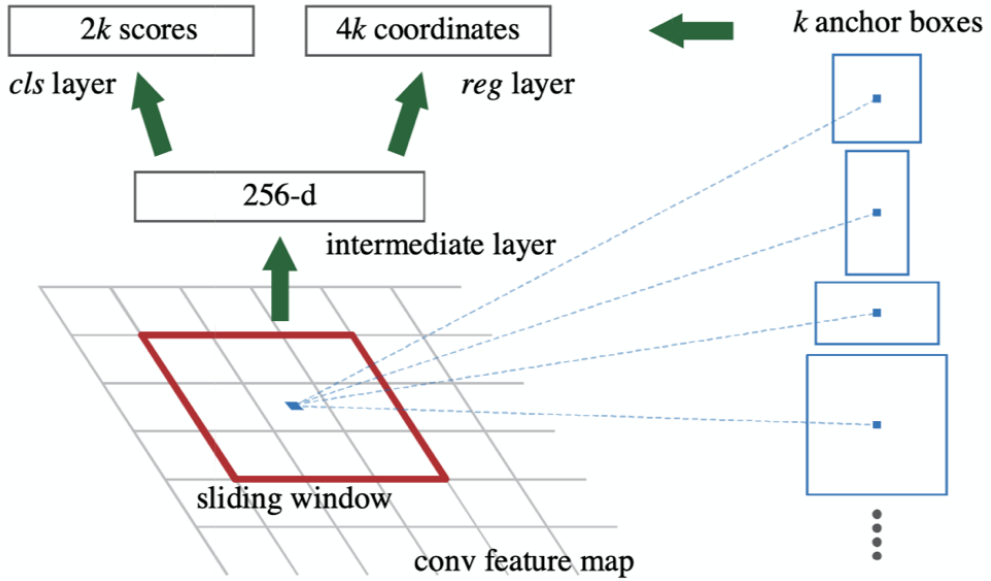




两阶段目标检测 - Faster R-CNN

Faster R-CNN的关键创新在于引入了Region Proposal Network（RPN），这个网络可以在特征图上快速生成候选区域（region proposals），然后将这些候选区域送入分类器和边界框回归器进行目标检测。通过共享卷积层来计算候选区域，Faster R-CNN实现了端到端的训练，从而提高了检测速度。

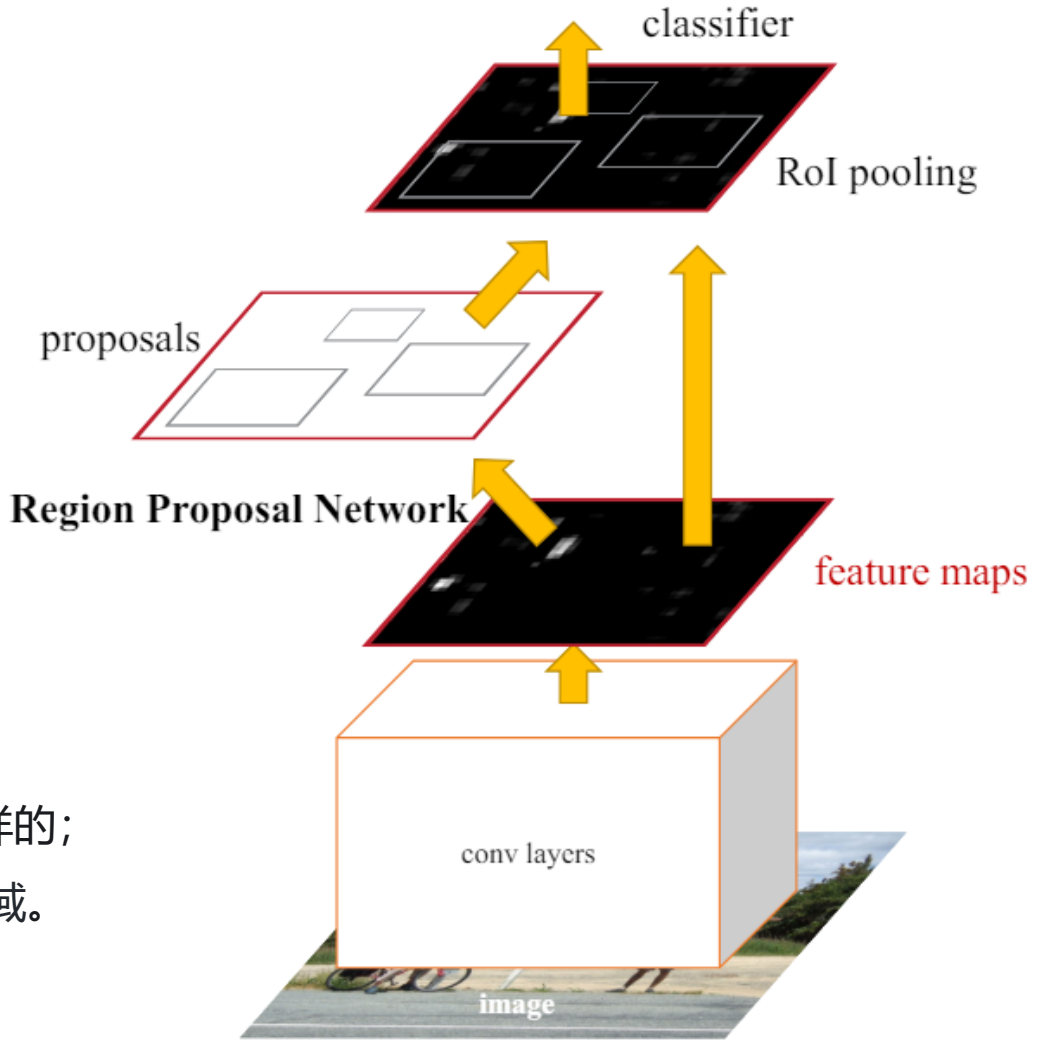




两阶段目标检测 - Faster R-CNN

模型结构：

- 输入图像通过卷积神经网络（如VGG、ResNet等）提取特征。
- 特征图通过Region Proposal Network生成候选区域。
- 候选区域经过RoI Pooling层进行特征提取。
- 最后，这些特征送入分类器和边界框回归器，得到目标类别和位置信息。



相对于Fast R-CNN 改进点：

- 真正做到了端到端（end-to-end）的目标检测；
- 提出网络参数共享，在不同任务中，用于特征提取的网络可以是一样的；
- 提出了RPN（Region Proposal Network），用来单独生成候选区域。


Thank
You