一 研读对象

论文:

<<Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks>>

二 研读笔记

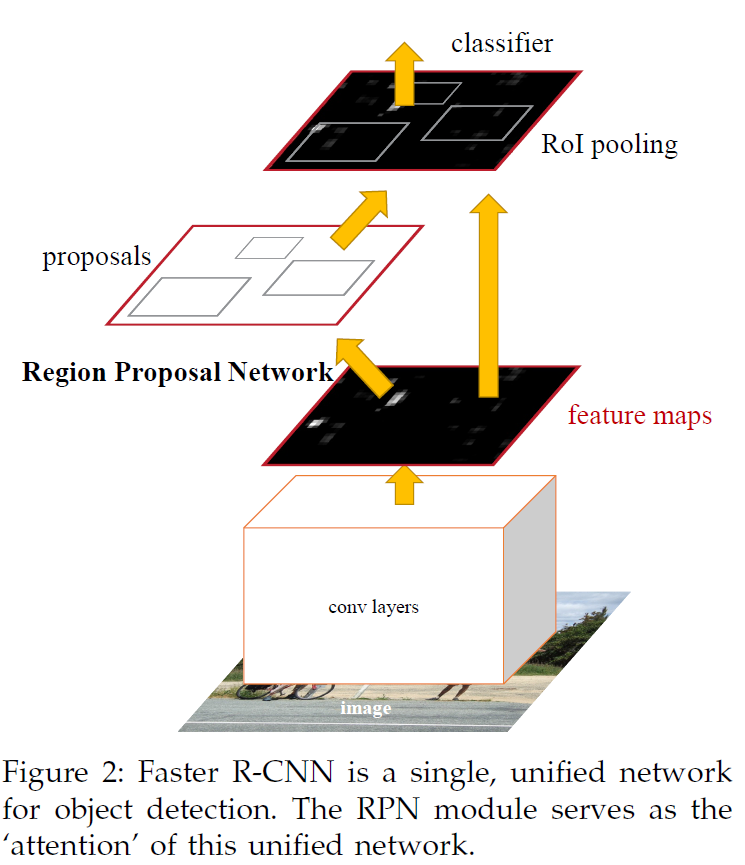
2.1 fr的基本结构

Fr 有两个模块.

其一,输出rpn框.

其二,做物体识别的推断.

但从代码上,从结构上看,两个模块组成统一的一个网络.如下图.



里边有分类(classifier),有proposals(rpn的部分).最终结果是分类.整体是端到端的.

Rpn会告诉cls在从哪儿去做分类.

2.2 Anchor

**Anchor综述**

Rpn的输入是image图.其输出是一系列的矩形框以及对应的一个得分.这个得分是描述该矩形框属于”物体对象”的程度(概率),区别于背景对象.

Rpn和detection基于同一组feature map.共享同一组conv layers.

|  |
| --- |
| 题外话:  Zeiler and Fergus(ZF model) 有5个共享conv layers.  Simonyan and Zisserman(VGG-16) 有13个共享conv layers. |

Rpn的输入是”共享conv layers”输出的feature map.在这组map上”滑动”一组network.具体的,这组网络以Map的nxn的窗口为输入.输出是低纬度的map, 这些低纬度map供cls和reg使用(cls和reg是兄弟关系sibling,他们也是全连接层).

Cls是说的矩形框是不是物体(“不是物体”指的是背景).

Reg是矩形框的位置,尺寸.

对于一个窗口有2+4个输出结果.

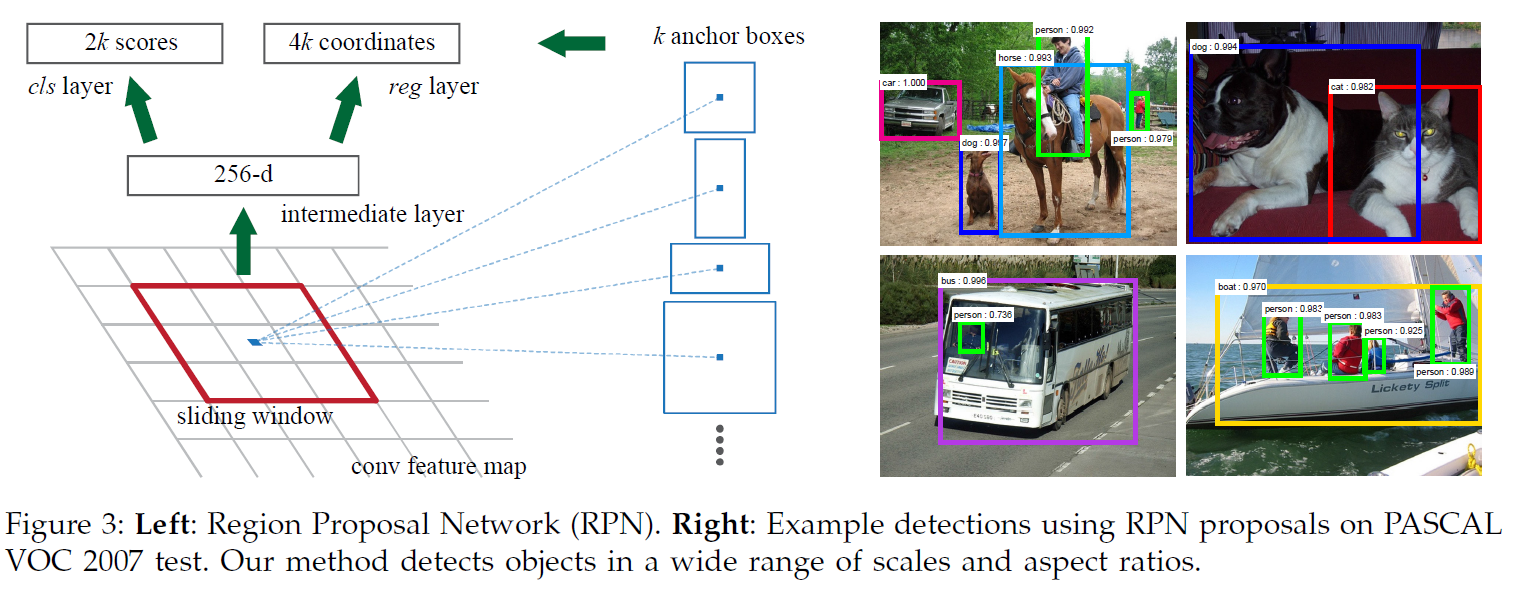
对于k个窗口有6k个结果.

一个”窗口”命名为anchor. Anchor含有scaler和ratio(详见vsd中关于anchor的说明部分).一个feature map的像素对应一组(scale,ratio的)anchor.(num\_scale \* num\_ratio, 以3x3为例,代码是3x4).一个像素对应k个anchor.

Map有WxH个”像素”,总共有WHk个anchors.

看到图中的:

1. Sliding window以及cls和reg layer.



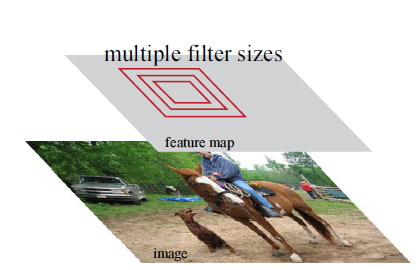
**Anchor相关点一: Anchor的相关网络具有平移不变性**

本质上由卷积+池化保证, 特别的有stride保证.

**Anchor相关点二: Anchor是多尺度的**

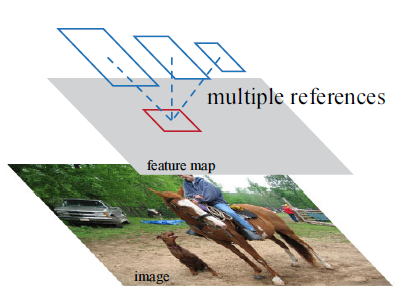
首先,多尺度预测有两种方法

1. 基于image/map的金字塔. 把图像resize到不同尺寸,相应的得到对应map. 它的缺点是非常耗时.
2. 对于同一组map,采用不同尺寸的”滑动窗口”(sliding windows), 不同的尺寸滑动窗口的模型会单独被训练,根据不同的filter.(5x7,和7x5).如下图, windows是不同尺寸的.组成的金字塔.此方法常与A方法联合使用.



特别的,引入anchors-pyramid.

Anchors框本身是多尺度的.基于此,会输出cls和reg.该方法的image和sliding window都不需要多尺寸.都是单一尺寸的.节省耗时.



2.3 loss function