**多目标检测算法(Faster R-CNN)心得**

1. **算法心得来源**：

对多目标检测算法Faster R-CNN 的心得来源学习华为云【ModelArts-Lab AI实战营】第五期：物体检测（ll），通过该次学习，从而了解**Faster R-CNN**网络结构及算法。

1. 网络结构组成:RPN+CNN+ROI三部分组成，其中：

RPN：为Region Proposal Network简写，中文意思为：“提取候选框”

CNN：卷积神经网络（深度学习，可参考之前学习的常用卷积神经网络学习算法：如VGG16模型）

ROI：为Region of Interest缩写,中文意思为：“特征图上的框（或特征区域）”

1. 算法思路：

* 候选区域生成
  + - 特征提取
* 分类，位置精修

该模型算法详情见以下章节

1. **目标检测（object detection）算法分类**：

根据目标检测发展和神经网络模型的发展，对目标检测算法主要分三个阶段，分别如下：

1. **传统的目标检测算法**：

Cascade + HOG/DPM + Haar/SVM以及上述方法的诸多改进、优化

1. **候选区域/框 + 深度学习分类**：通过提取候选区域，并对相应区域进行以深度学习方法为主的分类的方案，常用算法有如下：

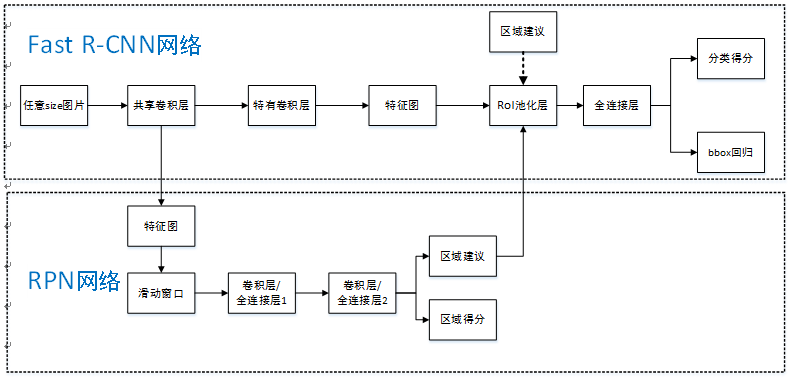
* R-CNN（Selective Search + CNN + SVM）
  + - SPP-net（ROI Pooling）
* Fast R-CNN（Selective Search + CNN + ROI）
  + - Faster R-CNN（RPN + CNN + ROI）
* R-FCN等系列方法

1. **基于深度学习的回归方法**：YOLO/SSD/DenseBox 等方法；以及最近出现的结合RNN算法的RRC detection；结合DPM的Deformable CNN等

在华为云【ModelArts-Lab AI实战营】第四期：物体检测（l）中采用YOLO模型；在华为云【ModelArts-Lab AI实战营】第五期：物体检测（ll）中采用Faster R-CNN模型，从目标检测发展历程算法：R-CNN、Fast R-CNN、Faster R-CNN可以看出，对候选区域采用高效先进的RPN网络替代了Select Search算法。

以下就以华为云【ModelArts-Lab AI实战营】第五期：物体检测（ll）中采用的算法Faster R-CNN重点做笔记整理和详细描述

1. **多目标检测Faster R-CNN算法详解**
2. **Faster R-CNN网络结构组成：**RPN+CNN+ROI,如下图：



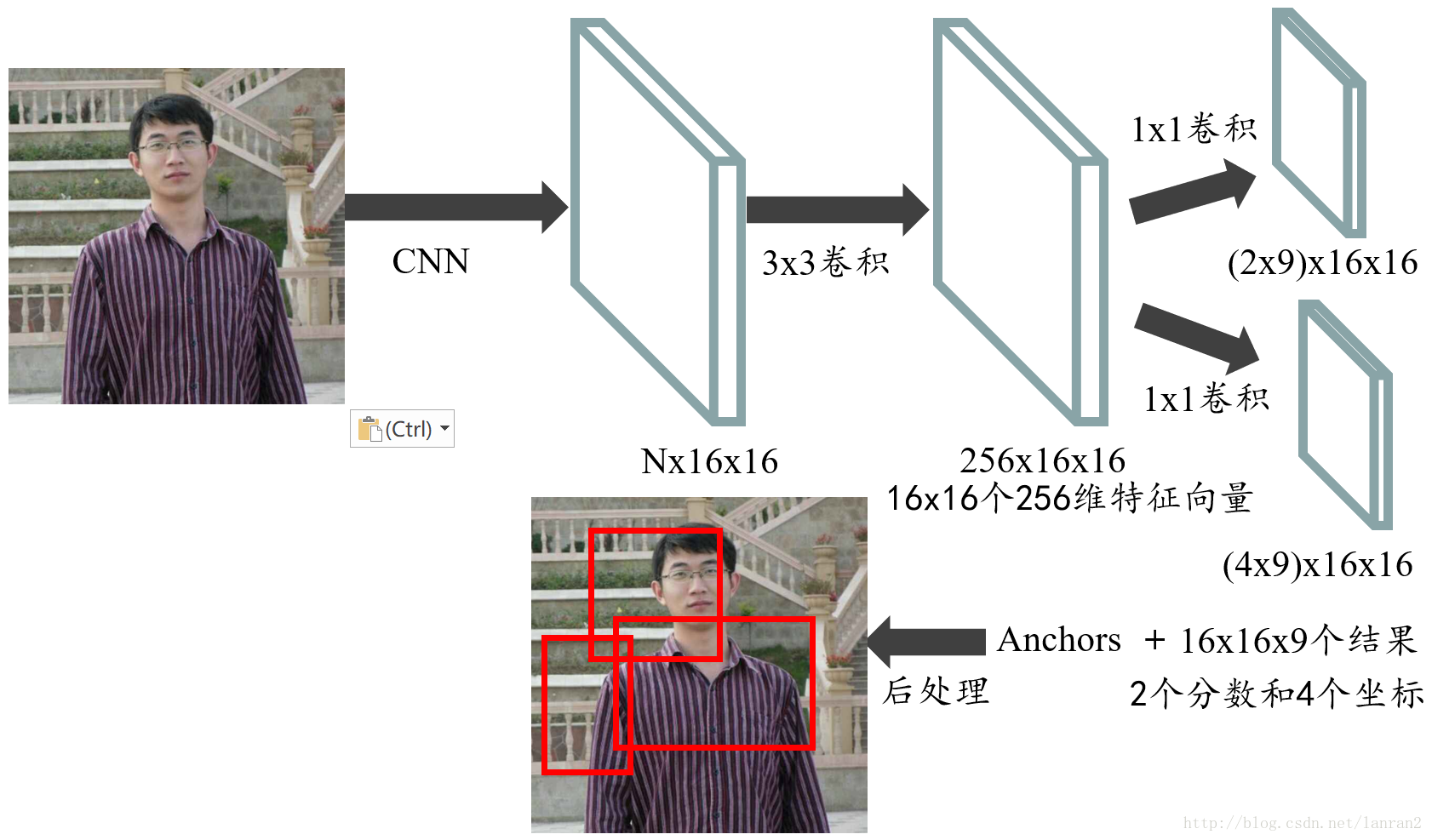
1. **Faster R-CNN网络思路如下**：

* 首先向CNN网络【VGG-16】输入任意大小图片M×NM×N
* 经过CNN网络前向传播至最后共享的卷积层，一方面得到供RPN网络输入的特征图，另一方面继续前向传播至特有卷积层，产生更高维特征图
* 供RPN网络输入的特征图经过RPN网络得到区域建议和区域得分，并对区域得分采用非极大值抑制【阈值为0.7】，输出其Top-N【文中为300】得分的区域建议给RoI池化层
* 第2步得到的高维特征图和第3步输出的区域建议同时输入RoI池化层，提取对应区域建议的特征
* 第4步得到的区域建议特征通过全连接层后，输出该区域的分类得分以及回归后的bounding-box

1. **RPN（区域生成网路）详解**：

由于在讲Faster R-CNN网络的时候，要使用到RPN,因此本节我们先讲讲RPN区域生成网络是如何来选择候选区域框的。RPN详细介绍请见（https://www.cnblogs.com/Terrypython/p/10584384.html），此处我只简单梳理RPN的处理流程：

首先通过一系列卷积得到公共特征图，假设他的大小是N x 16 x 16，然后我们进入RPN阶段，首先经过一个3 x 3的卷积，得到一个256 x 16 x 16的特征图，也可以看作16 x 16个256维特征向量，然后经过两次1 x 1的卷积，分别得到一个18 x 16 x 16的特征图，和一个36 x 16 x 16的特征图，也就是16 x 16 x 9个结果，每个结果包含2个分数和4个坐标，再结合预先定义的Anchors，经过后处理，就得到候选框；整个流程如图：

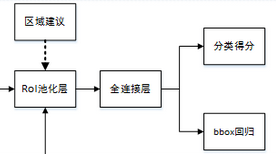


1. **RoI pooling层：**

在Faster R-CNN中提到了RoI pooling层，该层的作用是将不同大小尺寸的RoIs转换成统一固定长度大小的输出.

1. **分类与回归**：

通过RoI Pooling层我们已经得到所有候选区组成的特征向量，然后送入全连接层和softmax计算每个候选框具体属于哪个类别，输出类别的得分；同时再次利用框回归获得每个候选区相对实际位置的偏移量预测值，用于对候选框进行修正，得到更精确的目标检测框。



这里我们来看看全连接层，由于全连接层的参数w和b大小都是固定大小的，假设大小为49×26，那么输入向量的维度就要为Top−N×49，所以这就说明了RoI Pooling的重要性