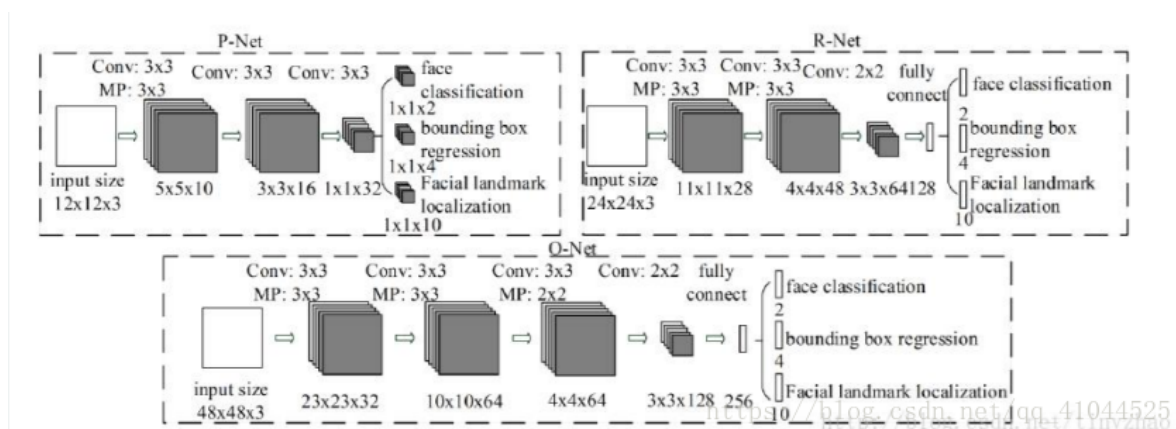


# MTCNN

MTCNN是个多任务级联CNN的人脸检测模型，由三个网络组成，我们把它叫做（P-Net、R-Net、O-Net）：



## P-Net:

这个网络的结构主要获取人脸区域的候选窗口以及边界框的回归向量。同时用这个边界框做回归，对候选窗口进行校准，然后通过非极大值抑制来合并高度重叠的候选框。

## R-Net:

该网络结构还是通过边界框回归和非极大值抑制来去掉一些错误区域。

## O-Net:

比R-Net层多了一层卷积层，所以处理更加精细，只是进行更多的监督，输出5个坐标。

## 训练:

对于三个部分，人脸的分类器（两个输出）：

$$L_i^{det} = -(y_i^{det} \log(p_i)) + (1 - y_i^{det})(1 - \log(p_i))$$
$$y_i^{det} \in \{0, 1\}$$

边界框回归（4个输出）：

$$L_i^{box} = ||y_i^{box} - \hat{y}_i^{box}||_2^2$$
$$y_i^{box} \in \mathbb{R}^4$$

y包含了左上角的x、y、长和宽。

地标定位（两个输出）：

$$\min \sum_{i=1}^N \sum_{j \in \{det, box, landmark\}} \alpha_h \beta_i^j L_i^j$$

## 非极大值抑制:

这里的非极大值抑制采用的是交并比这个数值，对于网络回归得到的多个人脸框，算出他们的交并比，根据一个阈值筛选出大于某个值得边框，再选出最大交并比的那个边框，然后留下与这个最大交并比边框的交并比小于某个阈值的边框，从而得到最终重叠出来的整个边框。

## **整体流程：**

多任务级联的卷积神经网络（MTCNN）的实现流程：首先是把图像经过金字塔处理，生成很多不同比例尺度的子图像，作为第一级卷积神经网络的输入，在训练的时候同时将有没有脸、脸的候选框的范式距离这样的多个任务整合到一个损失函数进行训练，然后我们在完成第一轮训练之后，根据生成的候选框位置进行一个非极大值抑制，得到最终的候选框，然后用这个候选框得到的图像作为第二级卷积神经网络的输入，同样我们需要做一个金字塔处理，重复第一级的操作，得到候选框后，再进行一个非极大值抑制，然后再传入第三级卷积神经网络完成训练。

三级网络的不同就是第二层比第一层多了一个全连接层，第三层比第二层多了一个卷积层。

## **池化层：**

下采样，特征降维、压缩数据，减小过拟合提高容错率

## **卷积层：**

局部感知，映射到隐藏特征空间

## **全连接层：**

将分布式特征映射到样本空间