MTCNN+MobileNet实现口罩检测: <https://www.bilibili.com/video/BV1i741177hd?p=1>

2.模型算法

通过比较成熟的人脸检测网络MTCNN检测出人脸的位置，把人脸截下来再传入新的Mobile Net神经网络中，判断是否佩戴了口罩。

->MTCNN->->MobileNet->

2.1 MTCNN

MTCNN，Multi-task convolutional neural network（多任务卷积神经网络），将人脸区域检测与人脸关键点检测放在了一起。该模型主要采用了三个级联的网络，采用候选框加分类器的思想，进行快速高效的人脸检测。**三个级联的网络分别是快速生成候选窗口的P-Net、进行高精度候选窗口过滤选择的R-Net和生成最终边界框与人脸关键点的O-Net。**和很多处理图像问题的卷积神经网络模型，该模型也用到了**图像金字塔、边框回归（binding-box regression）、非最大值抑制(NMS)**等技术。

实现流程：

->构造图像金字塔(缩为根2分之1=0.709)

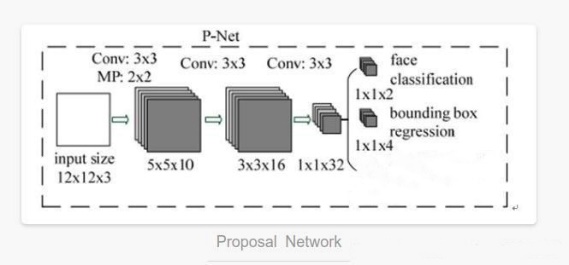
为了适应不同大小的人脸的进行检测。

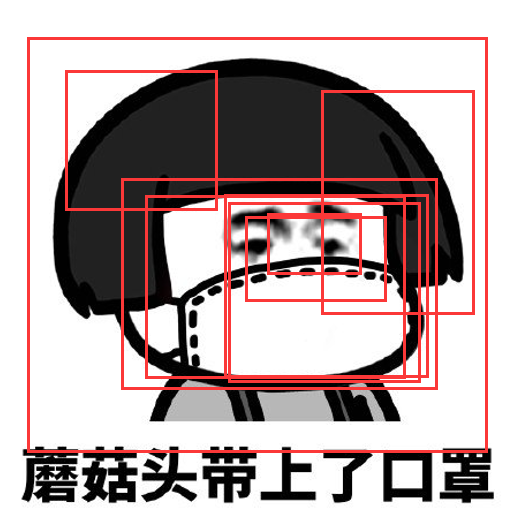


->构建P-Net

全称为Proposal Network，其基本的构造是一个全连接网络。**基本思想**是使用较为浅层、较为简单的CNN快速生成人脸候选窗口。对上一步构建完成的图像金字塔，通过一个全卷积网络（FCN）进行初步特征提取与标定边框，并进行边框回归（Bounding-Box Regression）调整窗口与非极大值抑制（NMS）进行大部分窗口的过滤。

P-Net是一个人脸区域的区域建议网络，该网络的将特征输入结果三个卷积层之后，通过一个人脸分类器判断该区域是否是人脸，同时使用边框回归来进行人脸区域的初步提议，该部分最终将输出很多张可能存在人脸的人脸区域，并将这些区域输入R-Net进行进一步处理。

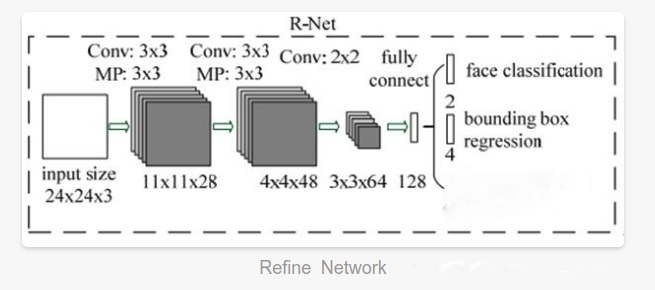


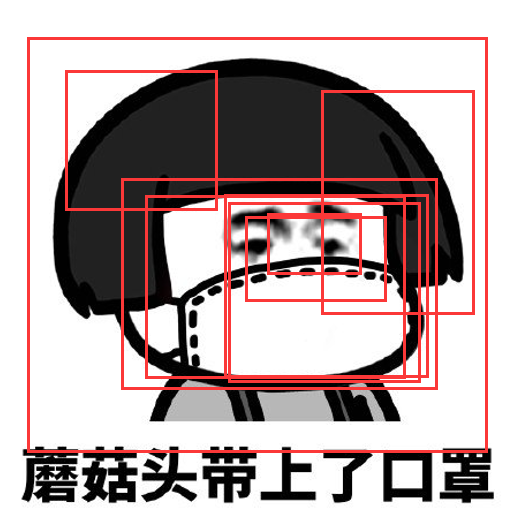


->构建R-Net

全称为Refine Network，其基本的构造是一个卷积神经网络，思想是使用一个相对于P-Net更复杂的网络结构来对P-Net生成的可能是人脸区域区域窗口进行进一步选择和调整，从而达到高精度过滤和人脸区域优化的效果。相对于第一层的P-Net来说，增加了一个全连接层，因此对于输入数据的筛选会更加严格。在图片经过P-Net后，会留下许多预测窗口，我们将所有的预测窗口送入R-Net，这个网络会滤除大量效果比较差的候选框，最后对选定的候选框进行Bounding-Box Regression和NMS进一步优化预测结果。

因为P-Net的输出只是具有一定可信度的可能的人脸区域，在这个网络中，将对输入进行细化选择，并且舍去大部分的错误输入，并再次使用边框回归进行人脸区域的边框回归，最后将输出较为可信的人脸区域，供O-Net使用。对比与P-Net使用全卷积输出的1x1x32的特征，R-Net使用在最后一个卷积层之后使用了一个128的全连接层，保留了更多的图像特征，准确度性能也优于P-Net。

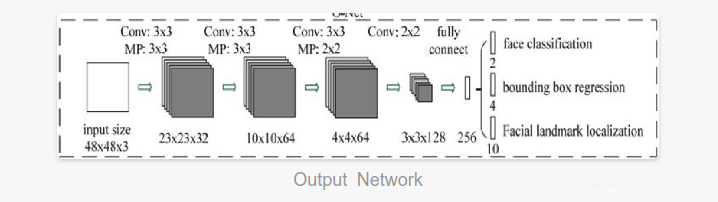




->构建O-Net

全称为Output Network，基本结构是一个较为复杂的卷积神经网络，相对于R-Net来说多了一个卷积层。O-Net的效果与R-Net的区别在于这一层结构会通过更多的监督来识别面部的区域，而且会对人的面部特征点进行回归，最终输出五个人脸面部特征点。

是一个更复杂的卷积网络，该网络的输入特征更多，在网络结构的最后同样是一个更大的256的全连接层，保留了更多的图像特征，同时再进行人脸判别、人脸区域边框回归和人脸特征定位，最终输出人脸区域的左上角坐标和右下角坐标与人脸区域的五个特征点。O-Net拥有特征更多的输入和更复杂的网络结构，也具有更好的性能，这一层的输出作为最终的网络模型输出。



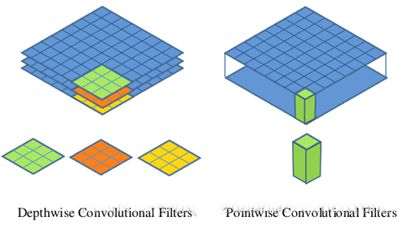


MTCNN为了兼顾性能和准确率，避免滑动窗口加分类器等传统思路带来的巨大的性能消耗，先使用小模型生成有一定可能性的目标区域候选框，然后在使用更复杂的模型进行细分类和更高精度的区域框回归，并且让这一步递归执行，以此思想构成三层网络，分别为P-Net、R-Net、O-Net，实现快速高效的人脸检测。在输入层使用图像金字塔进行初始图像的尺度变换，并使用P-Net生成大量的候选目标区域框，之后使用R-Net对这些目标区域框进行第一次精选和边框回归，排除大部分的负例，然后再用更复杂的、精度更高的网络O-Net对剩余的目标区域框进行判别和区域边框回归。

2.2 Mobile Net

Mobile Net属于直接设计小模型并进行训练，十分轻量，是Google最近提出的一种小巧而高效的CNN模型，其在accuracy和latency之间做了折中。

Mobile Net的基本单元是深度级可分离卷积（depthwise separable convolution），其实这种结构之前已经被使用在Inception模型中。深度级可分离卷积其实是一种可分解卷积操作，其可以分解为两个更小的操作：depthwise convolution和pointwise convolution



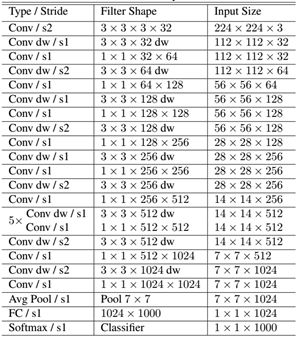
这里简单分析一下depthwise separable convolution在计算量上与标准卷积的差别**（轻量很多）**。假定输入特征图大小是，而输出特征图大小是  ，其中 是特征图的width和height，这是假定两者是相同的，而和指的是通道数（channels or depth）。这里也假定输入与输出特征图大小（width and height）是一致的。采用的卷积核大小是尽管是特例，但是不影响下面分析的一般性。对于标准的卷积 ，其计算量将是：

而对于depthwise convolution其计算量为： ，pointwise convolution计算量是： ，所以depthwise separable convolution总计算量是

可以比较depthwise separable convolution和标准卷积如下：

一般情况下 N 比较大，那么如果采用3x3卷积核的话，depthwise separable convolution相较标准卷积可以**降低大约9倍的计算量**。其实，后面会有对比，参数量也会减少很多。

Mobile Net网络结构



首先是一个3x3的标准卷积，然后后面就是堆积depthwise separable convolution，并且可以看到其中的部分depthwise convolution会通过strides=2进行down sampling。然后采用average pooling将feature变成1x1，根据预测类别大小加上全连接层，最后是一个softmax层。如果单独计算depthwise convolution和pointwise convolution，整个网络有28层（这里Avg Pool和Softmax不计算在内）。

参考资料：

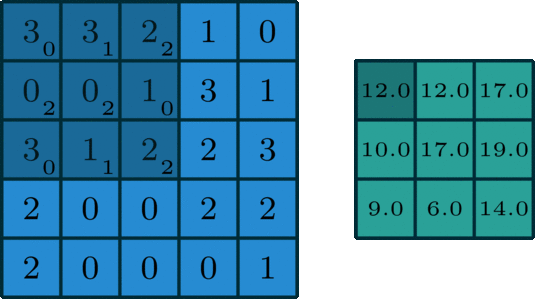
<https://blog.csdn.net/c20081052/article/details/80703896>

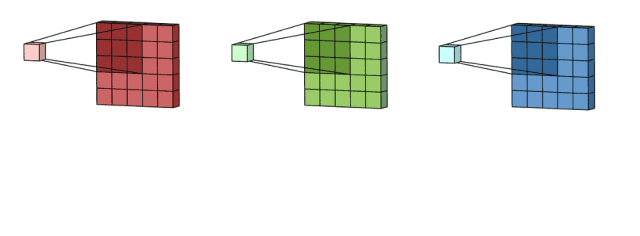
<https://www.bilibili.com/video/BV1U7411m7KL?p=2>

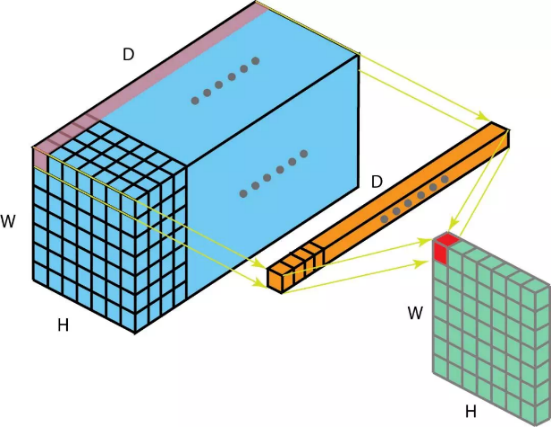
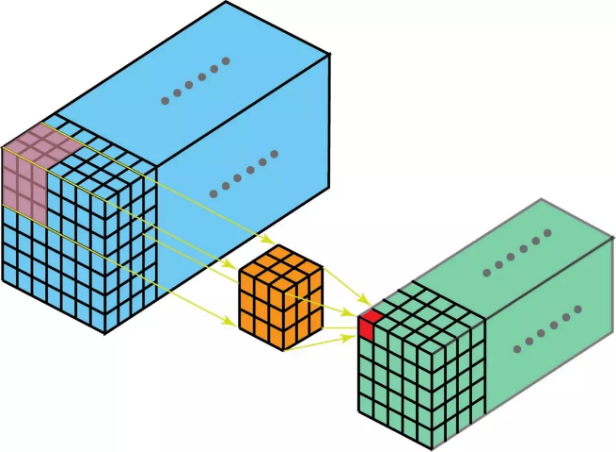
一些涉及的技术：

**卷积核：**

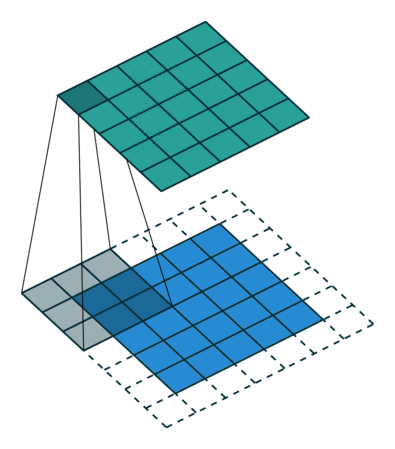
如何运作演示

****





下图是2D卷积，Kernel size=3，Stride=1，Padding=1：



**池化的理解：**<https://blog.csdn.net/sunflower_sara/article/details/81322048>

pooling的结果是使得特征减少，参数减少，但pooling的目的并不仅在于此。

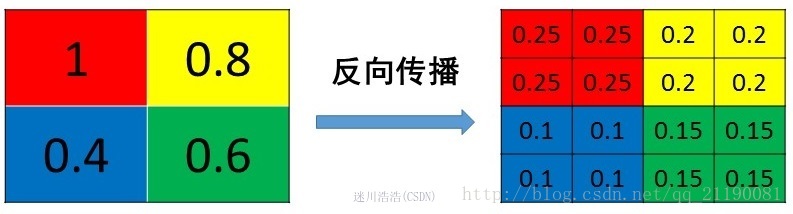
pooling目的是为了保持某种不变性（旋转、平移、伸缩等）

常用的有mean-pooling，max-pooling和Stochastic-pooling三种。

平均池化mean-pooling，即对邻域内特征点只求平均，

最大池化max-pooling，即对邻域内特征点取最大。

平均池化：



最大池化：



**FCN（全卷积网络）**

全卷积网络就是去除了传统卷积网络的全连接层，然后对其进行反卷积对最后一个卷积层（或者其他合适的卷积层）的feature map进行上采样，使其恢复到原有图像的尺寸（或者其他），并对反卷积图像的每个像素点都可以进行一个类别的预测，同时保留了原有图像的空间信息。

同时，在反卷积对图像进行操作的过程中，也可以通过提取其他卷积层的反卷积结果对最终图像进行预测，合适的选择会使得结果更好、更精细。

**IoU**

对于某个图像的子目标图像和对这个子目标图像进行标定的预测框，把最终标定的预测框与真实子图像的自然框（通常需要人工标定）的某种相关性叫做IOU（Intersection over Union），经常使用的标准为两个框的交叉面积与合并面积之和。

**Bounding-Box regression(边框回归)**

解决的问题：

当IOU小于某个值时，一种做法是直接将其对应的预测结果丢弃，而Bounding-Box regression的目的是对此预测窗口进行微调，使其接近真实值。

具体逻辑

在图像检测里面，子窗口一般使用四维向量(x,y,w,h)表示，代表着子窗口中心所对应的母图像坐标与自身宽高，目标是在前一步预测窗口对于真实窗口偏差过大的情况下，使得预测窗口经过某种变换得到更接近与真实值的窗口。

在实际使用之中，变换的输入输出按照具体算法给出的已经经过变换的结果和最终适合的结果的变换，可以理解为一个损失函数的线性回归。

**NMS（非极大值抑制）**

顾名思义，非极大值抑制就是抑制不是极大值的元素。在目标检测领域里面，可以使用该方法快速去掉重合度很高且标定相对不准确的预测框，但是这种方法对于重合的目标检测不友好。

**Soft-NMS**

对于优化重合目标检测的一种改进方法。核心在于在进行NMS的时候不直接删除被抑制的对象，而是降低其置信度。处理之后在最后统一一个置信度进行统一删除。

**PRelu**

在MTCNN中，卷积网络采用的激活函数是PRelu，带有参数的Relu，相对于Relu滤除负值的做法，PRule对负值进行了添加参数而不是直接滤除，这种做法会给算法带来更多的计算量和更多的过拟合的可能性，但是由于保留了更多的信息，也可能是训练结果拟合性能更好。

4.总结心得

现在已经有的人脸检测技术是针对脸部没有遮挡物的，所以该算法的效果可能会差一点，但是在光照充足的条件下情况就会好很多。而且我们的模型对于手部有遮挡物时可能会发生误判，所以加入了一些这样的图片再次进行训练，由于数据量害较小，虽然结果有一些改善，但还可以提升。