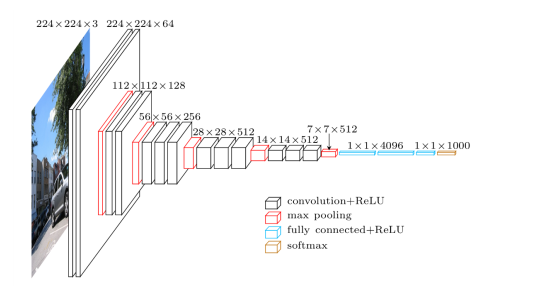
1.VGG论文给出了一个非常振奋人心的结论：卷积神经网络的深度增加和小卷积核的使用对网络的最终分类识别效果有很大的作用

2.VGG不仅在ILSVRC的分类和检测任务中取得了the state-of-the-art的精度，在其他数据集上也具有很好的推广能力

3.作者指出，虽然LRN(Local Response Normalisation)在AlexNet对最终结果起到了作用，但在VGG网络中没有效果，并且该操作会增加内存和计算，从而作者在更深的网络结构中，没有使用该操作。

4.VGG网络结构图（224\*224的RGB图像，唯一预处理，每个像素减去平均RGB值）



所有池化层2\*2的窗口，步幅为2，即缩小一半，非重叠池化

卷积核为3\*3 或者1\*1，步长为1

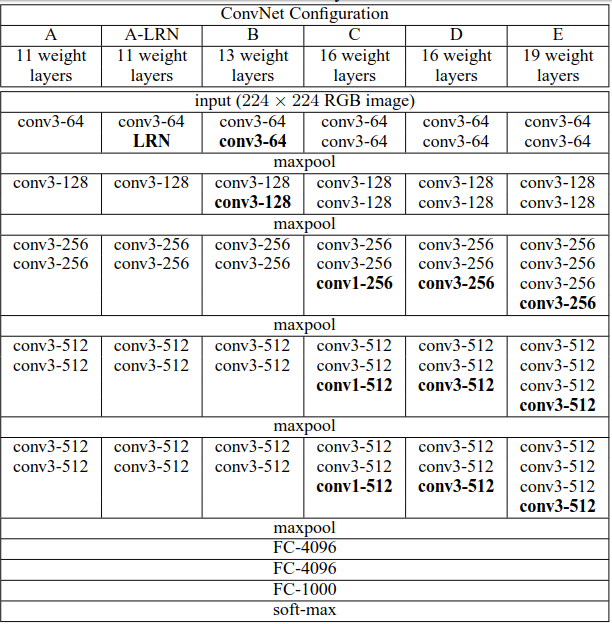
1x1卷积核：降维，增加非线性性

3x3卷积核：多个卷积核叠加，增加空间感受野，减少参数

激活函数ReLU

增加了对权重的正则化，5∗10−4||W||L2

对FC全连接层进行dropout正则化，dropout ratio = 0.5



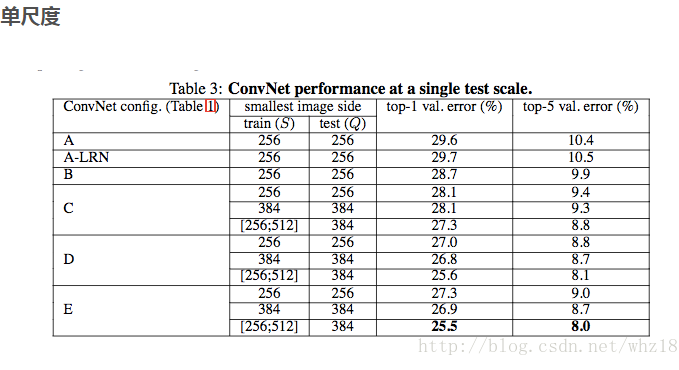
**训练输入：** 采用随机裁剪的方式，获取固定大小224x224的输入图像。并且采用了随机水平镜像和随机平移图像通道来丰富数据。

**Training image size:** 令S为图像的最小边，如果最小边S=224，则直接在图像上进行224x224区域随机裁剪，这时相当于裁剪后的图像能够几乎覆盖全部的图像信息；如果最小边S>>224，那么做完224x224区域随机裁剪后，每张裁剪图，只能覆盖原图的一小部分内容。 注：因为训练数据的输入为224x224，从而图像的最小边S，不应该小于224

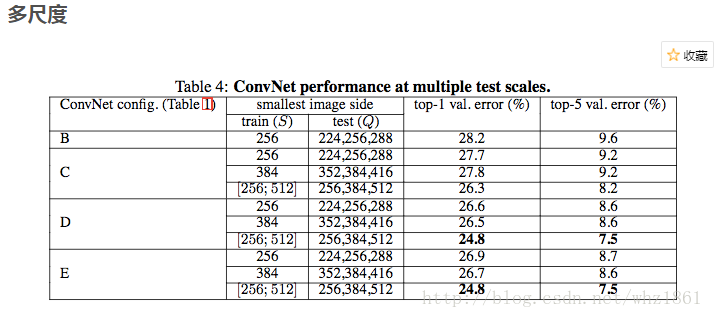
**数据生成方式：**首先对图像进行放缩变换，将图像的最小边缩放到S大小，然后

方法1: 在S=224和S=384的尺度下，对图像进行224x224区域随机裁剪

方法2: 令S随机的在[Smin,Smax]区间内值，放缩完图像后，再进行随机裁剪（其中Smin=256,Smax=512Smin=256,Smax=512）



* 模型E（VGG19）的效果最好，即网络越深，效果越好
* 同一种模型，随机scale jittering的效果好于固定S大小的256，384两种尺度，即scale jittering数据增强能更准确的提取图像多尺度信息



* 对比单尺度预测，多尺度综合预测，能够提升预测的精度
* 同单尺度预测，多尺度预测也证明了scale jittering的作用