

## VGGNet Notes

### 1. Architecture

- 1.1. 这篇论文主要关注 CNN 的深度对分类效果的影响。
- 1.2. 训练图片的输入：224\*224 RGB 图片，数据预处理只用了减去 RGB 平均值。
- 1.3. 网络中负责提取特征的卷积核都是 3\*3 大小的，这个大小的卷积核是能捕捉到上下左右概念的最小卷积核。  
Stride=1, Padding=1, 这样可以使卷积操作之后空间分辨率不变。Max Pooling 2\*2, stride=2。
- 1.4. 局部响应归一化没什么用。

### 2. Network Configuration

Table 1: **ConvNet configurations** (shown in columns). The depth of the configurations increases from the left (A) to the right (E), as more layers are added (the added layers are shown in bold). The convolutional layer parameters are denoted as “conv(receptive field size)-(number of channels)”. The ReLU activation function is not shown for brevity.

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

- 
- 2.1. 两个  $3 \times 3$  的卷积核堆叠，感受野和一个  $5 \times 5$  卷积核相同，但是增加了一个非线性函数，三个  $3 \times 3$  卷积核堆叠相当于一个  $7 \times 7$  卷积核。
  - 2.2. 使用较小的卷积核，可以显著降低参数量，如：假设输入、输出通道数量都是  $C$ ，使用三个  $3 \times 3$  卷积核的参数量是  $3 \times 3^2 \times C^2$ （输入层  $C$  个通道，输出层  $C$  个特征图），而使用一个  $7 \times 7$  卷积核的参数量是  $7^2 C^2$ 。
  - 2.3. 这里效仿 NIN，使用  $1 \times 1$  的卷积核增加非线性。

### 3. 训练过程

- 3.1. batch size 256, momentum 0.9,  $L_2$  weight decay  $5 \times 10^{-4}$ , 前两个全连接层 dropout 0.5
- 3.2. 训练图片的获取。令  $S$  等于原始图片 resize 后较短边的长度，训练所用图片是从 resize 的图片中裁剪  $224 \times 224$  大小的图片。有两种设置  $S$  的方法，第一种：固定  $S$ ，论文中使用了  $S=256$ 、 $384$  两种尺寸分别训练了模型；第二种：随机从  $[S_{\min}, S_{\max}]$  中采样  $S$ ， $S_{\min}=256$ ， $S_{\max}=512$ 。

### 4. 测试过程

- 4.1. 这个要参考 OverFeat 这篇论文，作者说的 dense 什么意思没懂。

### 5. 实验

- 5.1. 两种测试方法：dense、multi-crop

这两种方法互补，因为结合使用时，效果比任何一个都要好  
 Table 5: **ConvNet evaluation techniques comparison.** In all experiments the training scale  $S$  was sampled from  $[256; 512]$ , and three test scales  $Q$  were considered:  $\{256, 384, 512\}$ .

ConvNet config. (Table 1)	Evaluation method	top-1 val. error (%)	top-5 val. error (%)
D	dense	24.8	7.5
	multi-crop	24.6	7.5
	multi-crop & dense	<b>24.4</b>	<b>7.2</b>
E	dense	24.8	7.5
	multi-crop	24.6	7.4
	multi-crop & dense	<b>24.4</b>	<b>7.1</b>

5.2. dense 这个方法出自 OverFeat，先把图片 resize 成  $Q$ ，  
 然后应用 FCN 获得最后的 feature map， $Q$  不必等于  $S$ 。  
 multi-crop 出自 GoogLeNet，看到再说