Alex在2012年提出的alexnet网络结构模型引爆了神经网络的应用热潮，并赢得了2012届图像识别大赛的冠军，使得CNN成为在图像分类上的核心算法模型。

**接下来本文对该网络配置结构中各个层进行详细的解读（训练阶段）：**



[227x227x3]Input

[55x55x96] Conv1:96 11x11×3 filters at stride 4, pad 0

[27x27x96] Max Pool1:3X3 filters at stride 2

[27x27x256]Conv2:256 5x5×96 filters at stride 1, pad 2

[13x13x256]Max Pool2:3X3 filters at stride 2

[13x13x384]Conv3:384 3x3×256 filters at stride 1, pad 1

[13x13x384]Conv4:384 3x3×384 filters at stride 1, pad 1

[13x13x256]Conv5:256 3x3×384 filters at stride 1, pad 1

[6x6x256]Max Pool3:3X3 filters at stride 2

[4096]FC6:4096 6×6×256 neurons

[4096]FC7:4096 1×1×4096 neurons

[1000]FC8:1000 1×1×4096 neurons(class scores)

 1. **conv1阶段DFD（data flow diagram）：**



第一层输入数据为原始的227\*227\*3的图像，这个图像被11\*11\*3的卷积核进行卷积运算，卷积核对原始图像的每次卷积都生成一个新的像素。卷积核沿原始图像的x轴方向和y轴方向两个方向移动，移动的步长是4个像素。因此，卷积核在移动的过程中会生成(227-11)/4+1=55个像素(227个像素减去11，正好是54，即生成54个像素，再加上被减去的11也对应生成一个像素)，行和列的55\*55个像素形成对原始图像卷积之后的像素层。共有96个卷积核，会生成55\*55\*96个卷积后的像素层。96个卷积核分成2组，每组48个卷积核。对应生成2组55\*55\*48的卷积后的像素层数据。这些像素层经过relu1单元的处理，生成激活像素层，尺寸仍为2组55\*55\*48的像素层数据。

这些像素层经过pool运算(池化运算)的处理，池化运算的尺度为3\*3，运算的步长为2，则池化后图像的尺寸为(55-3)/2+1=27。 即池化后像素的规模为27\*27\*96；然后经过归一化处理，归一化运算的尺度为5\*5；第一卷积层运算结束后形成的像素层的规模为27\*27\*96。分别对应96个卷积核所运算形成。这96层像素层分为2组，每组48个像素层，每组在一个独立的GPU上进行运算。

反向传播时，每个卷积核对应一个偏差值。即第一层的96个卷积核对应上层输入的96个偏差值。

2.**conv2阶段DFD（data flow diagram）：**



第二层输入数据为第一层输出的27\*27\*96的像素层，为便于后续处理，每幅像素层的左右两边和上下两边都要填充2个像素；27\*27\*96的像素数据分成27\*27\*48的两组像素数据，两组数据分别再两个不同的GPU中进行运算。每组像素数据被5\*5\*48的卷积核进行卷积运算，卷积核对每组数据的每次卷积都生成一个新的像素。卷积核沿原始图像的x轴方向和y轴方向两个方向移动，移动的步长是1个像素。因此，卷积核在移动的过程中会生成(27-5+2\*2)/1+1=27个像素。(27个像素减去5，正好是22，在加上上下、左右各填充的2个像素，即生成26个像素，再加上被减去的5也对应生成一个像素)，行和列的27\*27个像素形成对原始图像卷积之后的像素层。共有256个5\*5\*48卷积核；这256个卷积核分成两组，每组针对一个GPU中的27\*27\*48的像素进行卷积运算。会生成两组27\*27\*128个卷积后的像素层。这些像素层经过relu2单元的处理，生成激活像素层，尺寸仍为两组27\*27\*128的像素层。

这些像素层经过pool运算(池化运算)的处理，池化运算的尺度为3\*3，运算的步长为2，则池化后图像的尺寸为(27-3)/2+1=13。 即池化后像素的规模为2组13\*13\*128的像素层；然后经过归一化处理，归一化运算的尺度为5\*5；第二卷积层运算结束后形成的像素层的规模为2组13\*13\*128的像素层。分别对应2组128个卷积核所运算形成。每组在一个GPU上进行运算。即共256个卷积核，共2个GPU进行运算。

反向传播时，每个卷积核对应一个偏差值。即第一层的96个卷积核对应上层输入的256个偏差值。

 3. **conv3阶段DFD（data flow diagram）：**



第三层输入数据为第二层输出的2组13\*13\*128的像素层；为便于后续处理，每幅像素层的左右两边和上下两边都要填充1个像素；2组像素层数据都被送至2个不同的GPU中进行运算。每个GPU中都有192个卷积核，每个卷积核的尺寸是3\*3\*256。因此，每个GPU中的卷积核都能对2组13\*13\*128的像素层的所有数据进行卷积运算。卷积核对每组数据的每次卷积都生成一个新的像素。卷积核沿像素层数据的x轴方向和y轴方向两个方向移动，移动的步长是1个像素。因此，运算后的卷积核的尺寸为(13-3+1\*2)/1+1=13（13个像素减去3，正好是10，在加上上下、左右各填充的1个像素，即生成12个像素，再加上被减去的3也对应生成一个像素），每个GPU中共13\*13\*192个卷积核。2个GPU中共13\*13\*384个卷积后的像素层。这些像素层经过relu3单元的处理，生成激活像素层，尺寸仍为2组13\*13\*192像素层，共13\*13\*384个像素层。

 4.**conv4阶段DFD（data flow diagram）：**



第四层输入数据为第三层输出的2组13\*13\*192的像素层；为便于后续处理，每幅像素层的左右两边和上下两边都要填充1个像素；2组像素层数据都被送至2个不同的GPU中进行运算。每个GPU中都有192个卷积核，每个卷积核的尺寸是3\*3\*192。因此，每个GPU中的卷积核能对1组13\*13\*192的像素层的数据进行卷积运算。卷积核对每组数据的每次卷积都生成一个新的像素。卷积核沿像素层数据的x轴方向和y轴方向两个方向移动，移动的步长是1个像素。因此，运算后的卷积核的尺寸为(13-3+1\*2)/1+1=13（13个像素减去3，正好是10，在加上上下、左右各填充的1个像素，即生成12个像素，再加上被减去的3也对应生成一个像素），每个GPU中共13\*13\*192个卷积核。2个GPU中共13\*13\*384个卷积后的像素层。这些像素层经过relu4单元的处理，生成激活像素层，尺寸仍为2组13\*13\*192像素层，共13\*13\*384个像素层。

 5. conv5阶段DFD（data flow diagram）：



第五层输入数据为第四层输出的2组13\*13\*192的像素层；为便于后续处理，每幅像素层的左右两边和上下两边都要填充1个像素；2组像素层数据都被送至2个不同的GPU中进行运算。每个GPU中都有128个卷积核，每个卷积核的尺寸是3\*3\*192。因此，每个GPU中的卷积核能对1组13\*13\*192的像素层的数据进行卷积运算。卷积核对每组数据的每次卷积都生成一个新的像素。卷积核沿像素层数据的x轴方向和y轴方向两个方向移动，移动的步长是1个像素。因此，运算后的卷积核的尺寸为(13-3+1\*2)/1+1=13（13个像素减去3，正好是10，在加上上下、左右各填充的1个像素，即生成12个像素，再加上被减去的3也对应生成一个像素），每个GPU中共13\*13\*128个卷积核。2个GPU中共13\*13\*256个卷积后的像素层。这些像素层经过relu5单元的处理，生成激活像素层，尺寸仍为2组13\*13\*128像素层，共13\*13\*256个像素层。

2组13\*13\*128像素层分别在2个不同GPU中进行池化(pool)运算处理。池化运算的尺度为3\*3，运算的步长为2，则池化后图像的尺寸为(13-3)/2+1=6。 即池化后像素的规模为两组6\*6\*128的像素层数据，共6\*6\*256规模的像素层数据。

 6. fc6阶段DFD（data flow diagram）：



第六层输入数据的尺寸是6\*6\*256，采用6\*6\*256尺寸的滤波器对第六层的输入数据进行卷积运算；每个6\*6\*256尺寸的滤波器对第六层的输入数据进行卷积运算生成一个运算结果，通过一个神经元输出这个运算结果；共有4096个6\*6\*256尺寸的滤波器对输入数据进行卷积运算，通过4096个神经元输出运算结果；这4096个运算结果通过relu激活函数生成4096个值；并通过drop运算后输出4096个本层的输出结果值。

由于第六层的运算过程中，采用的滤波器的尺寸(6\*6\*256)与待处理的feature map的尺寸(6\*6\*256)相同，即滤波器中的每个系数只与feature map中的一个像素值相乘；而其它卷积层中，每个滤波器的系数都会与多个feature map中像素值相乘；因此，将第六层称为全连接层。

第五层输出的6\*6\*256规模的像素层数据与第六层的4096个神经元进行全连接，然后经由relu6进行处理后生成4096个数据，再经过dropout6处理后输出4096个数据。

 7. fc7阶段DFD（data flow diagram）：



第六层输出的4096个数据与第七层的4096个神经元进行全连接，然后经由relu7进行处理后生成4096个数据，再经过dropout7处理后输出4096个数据。

 8. fc8阶段DFD（data flow diagram）：



第七层输出的4096个数据与第八层的1000个神经元进行全连接，经过训练后输出被训练的数值。