《ImageNet Classification With Deep Convolutional Neural Networks》

**主要架构：卷积神经网络（CNNs）：**

最终的神经网络包含5个卷积层和3个全连接层，并且在论文中表明这个深度似乎是很重要的：去掉任意一个卷积层（每一层包含的参数个数不超过整个模型参数个数的1%）都会导致更差的表现。

**数据集：**

ImageNet数据集（包含大概22000种共1500多万的带标签的高清图片）。

**模型体系结构：**

（1）非线性函数：；（即激活函数）

修正非线性单元（Rectified Linear Units ()），相比传统的饱和非线性函数，如，它能够有效加快大型模型的训练速度。

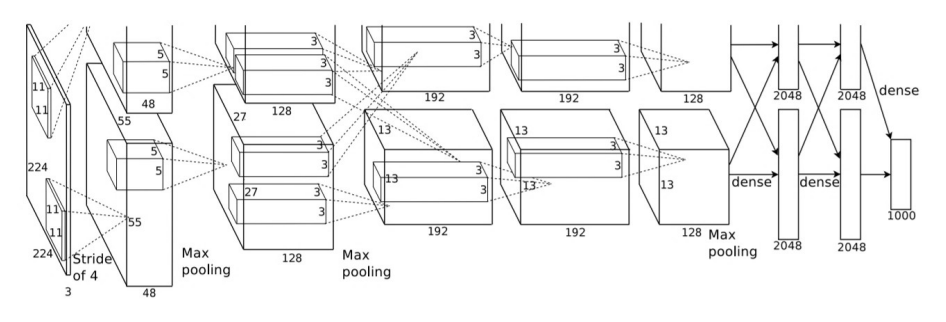
（2）多GPU并行训练；

（3）局部响应归一化；

（4）重叠池化；

CNN中的池化层总结了同一核映射上相邻神经元的输出。而一般来说，相邻池化单元总结的邻近关系是不重叠的，但本模型采用了重叠池化单元的方法。

（5）总体结构：



CNN体系结构示意图，明确显示了两个GPU之间的职责划分。一个GPU运行图中顶部的层次部分，而另一个GPU运行图中底部的层次部分。此外，GPU之间仅在某些层互相通信。

**两种减少过拟合的主要方式：**

1. 数据增强；

图像变换（裁剪）和水平翻转，改变训练图像的RGB通道的强度，采用PCA方法。

（2）随机失活；（这在当时是最新的正则化方法）

一种非常有效的避免过拟合的方法，但会增加收敛所需要的迭代次数。

**模型学习细节：**

（1）采用随机梯度下降法（）；

（2）（动量梯度下降），（即L2正则项）

（3）权重的更新规则：（）

（4）权重初始化；

使用均值为0，标准差为0.01的高斯分布对每一层的权重进行初始化。使用常数1来初始化第二、四、五卷积层和全连接层的神经元的偏置项，使用常数0来初始化其余层的神经元偏置项。（5）学习率的设置；

对所有层采用相同的学习率，并且在训练过程中动态调整学习率。

**训练结果：**

模型的误差率在当时远低于其他训练模型在图片分类问题中的误差率，模型非常有效。同时，本论文也首次表明了大型深度卷积神经网络在一个非常具有挑战性的数据集上使用纯粹的监督学习，能够达到破纪录的效果。