AlexNet

**计算机视觉可以认为是在根据不同的问题抽取不同的特征。**

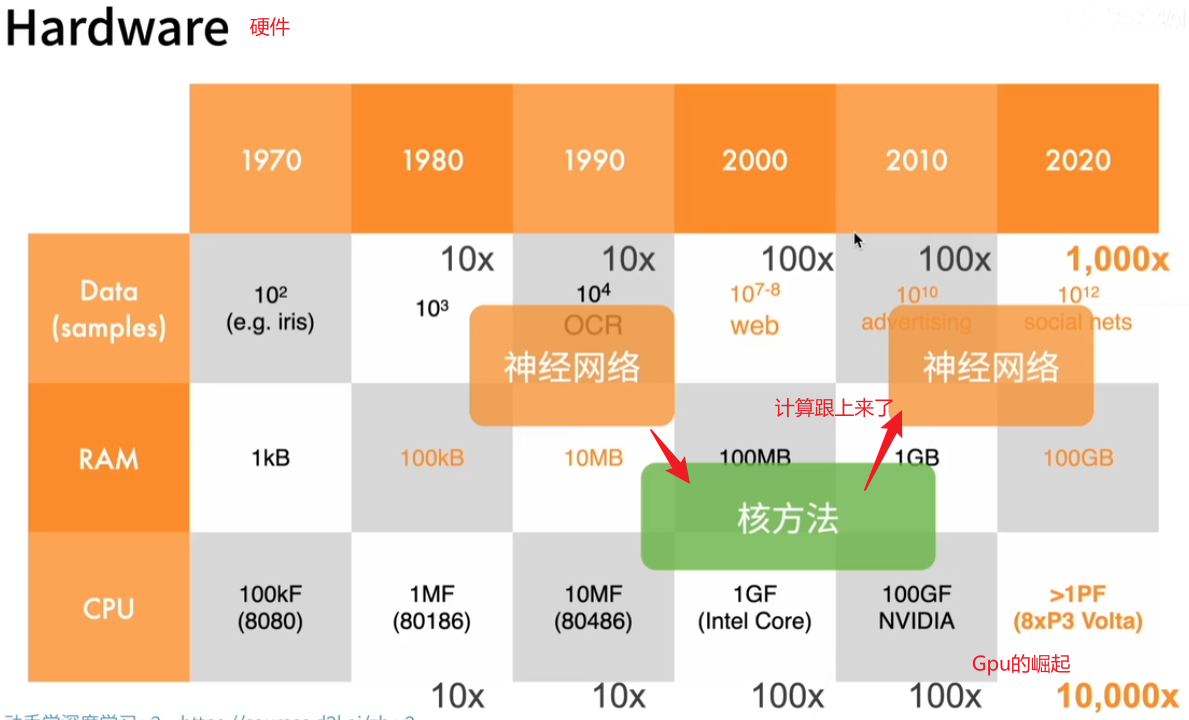
使用深度卷积神经网络进行 ImageNet 分类：

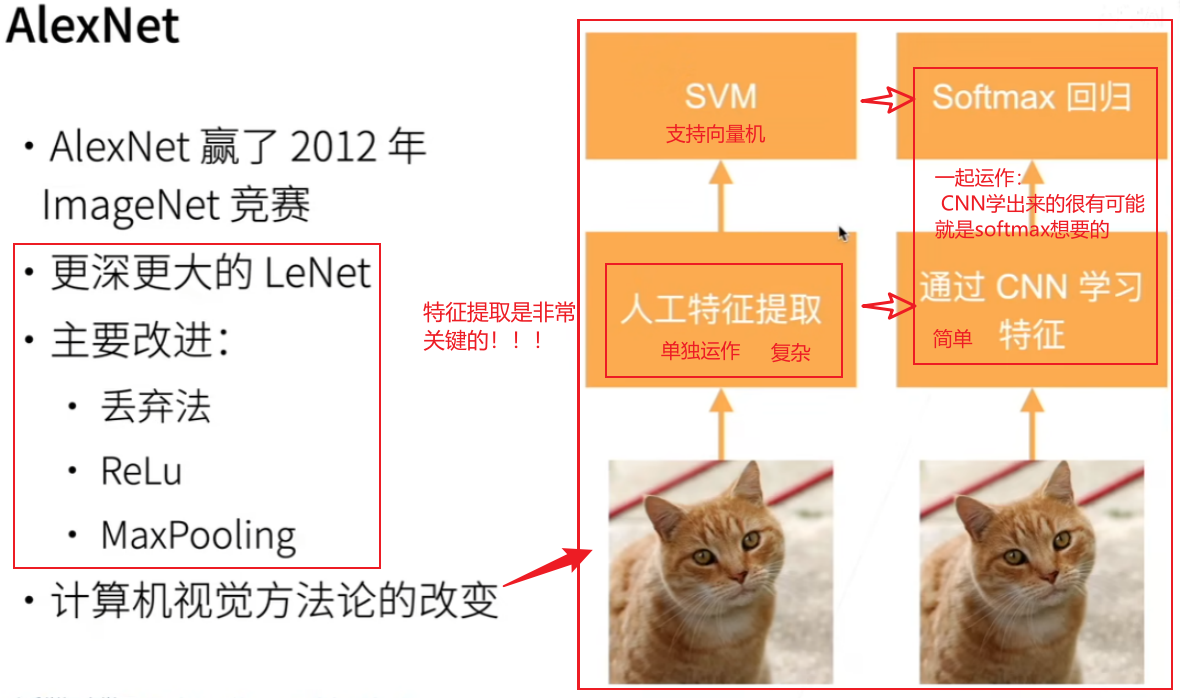
1.大型卷积神经网络AlexNet，实现了低错误率。

2.该网络由5个卷积层和3个全连接层组成。

3.使用了GPU加快了训练速度。

4.开发了“dropout”正则化手段减少过拟合。





一、 简介

为了更强大的模型，大型的数据集是需要的，如ImageNet。

GPU能够与2D卷积实现结合，可以促进CNN训练。比CPU快几个数量级。  
卷积神经网络中的计算瓶颈：卷**积和矩阵乘法，都是可以在硬件上并行化的操作。**

如果有更快的GPU和更大的数据集，结果就可以得到改善。

二、数据集

**使用ImageNet的子集，每个类别包含大约1000个图像。总共大约有120万张训练图像、5五万张验证图像和15万张测试图像。**

**ImageNet是由可变分辨率图像组成，这里将图像下采样到固定分辨率256 × 256。**

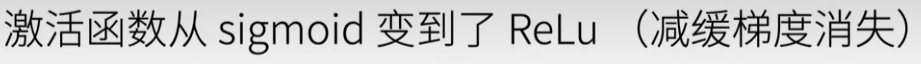
(ImageNet数据集由斯坦福教授李飞飞小组的研究人员开发，利用谷歌图像搜索对每一类图像进行预筛选，并利用亚马逊众包来标注每张图片的相关类别。其有100万个样本中训练模型，以区分1000个不同类别的对象。)



三、架构

3.1 ReLU非线性

使用 ReLU 的深度卷积神经网络的训练速度比使用 tanh 单元的深度卷积神经网络快几倍。



3.2 多GPU上的训练

GPU适合并行训练，将网络分布在两个GPU上。

3.3 局部响应标准化

3.4 重叠池化

间隔s个像素<池化单元位置为中心的大小为z × z 的邻域。

3.5 整体架构

输入图像：224×224×3

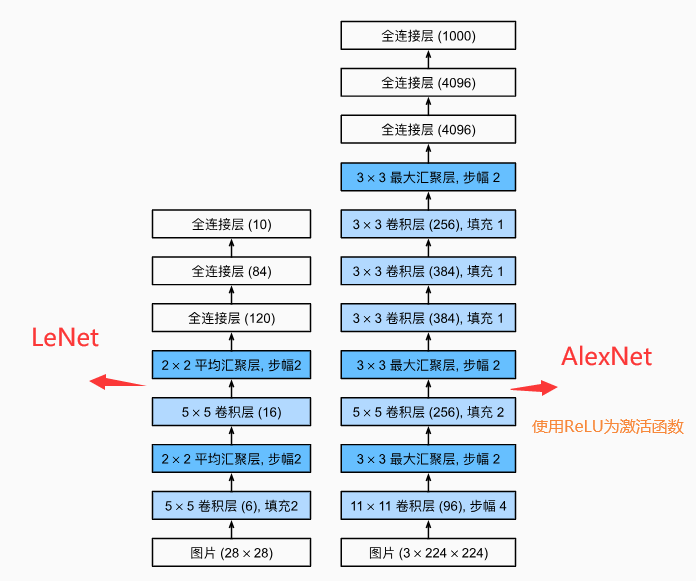
第一个卷积层 ：96 个大小为 11×11×3 的核，步幅4

第二个卷积层：256 个大小为 5 × 5 × 48核

第三个卷积层：384 个大小为 3 × 3 × 256 的内核

第四个卷积层： 384 个大小为 3 × 3 × 192 的内核

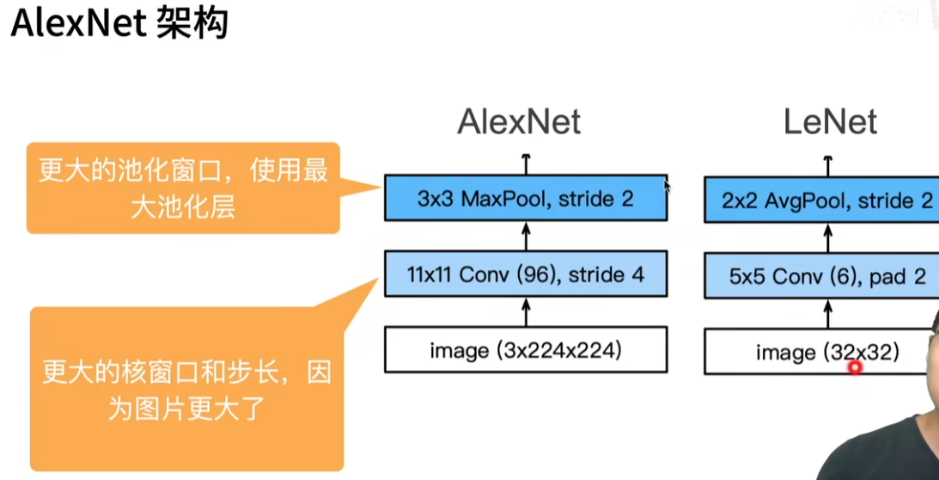
第五个卷积层： 256 个大小为 3 × 3 × 192 的内核

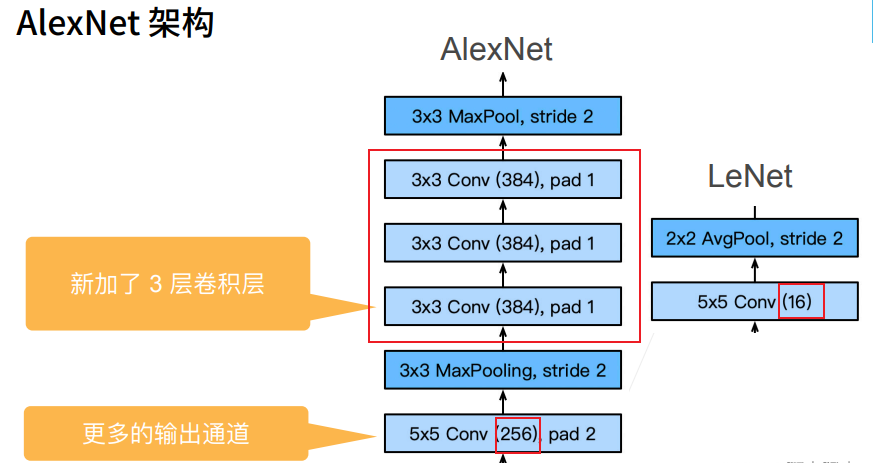
第一个全连接层：4096 个神经元

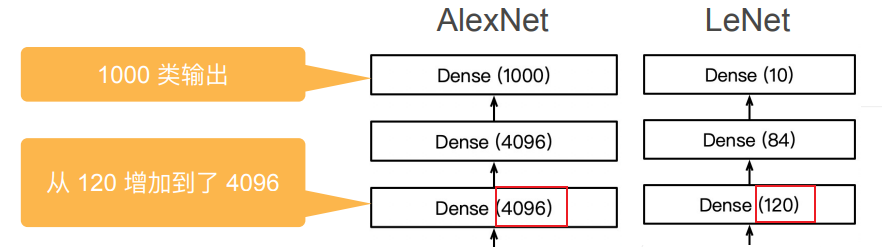
第二个全连接层：4096 个神经元

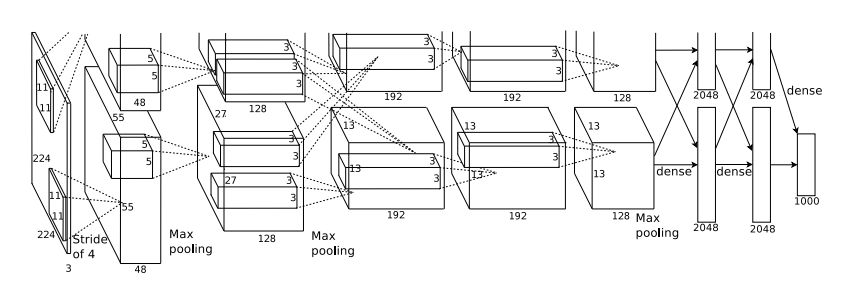
第三个全连接层：1000个神经元

最后输出到softmax









四、 减少过拟合

4.1 数据增强（后续课程学）

图像平移和水平反射，将训练集的大小增加了 2048 倍，减少了过拟合。

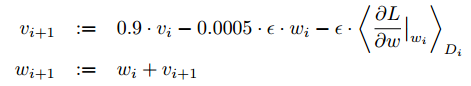
改变训练图像中 RGB 通道的强度，将错误率降低了1%。

4.2 Dropout丢弃法

以 0.5 的概率将每个隐藏神经元的输出设置为零，减少了过拟合。

五、学习细节

5.1使用随机梯度下降来训练模型，减少了模型的训练误差



5.2 使用标准差为0.01的零均值高斯分布初始化每层权重。

用常量1初始化第二、第四、第五卷积层和全连接隐藏层的神经元偏差。

用常量0初始化剩余层的神经元偏差。

5.3对所有层使用相同的学习率，当验证错误率不随当前学习率提高，将学习率除以10。学习率初始化为0.01并且终止前减少了三倍。

六、结果

错误率相较于之前的Top-1和Top-5得到明显降低

删除单个卷积层，网络性能就会下降，所以深度对于实现图像分类很重要。

6.1 定性评价

**1. 学习表征**

**特征本身应该被学习**

1．在合理地复杂性前提下，特征应该由多个共同学习的神经网络层组成，每个层都有可学习的参数。

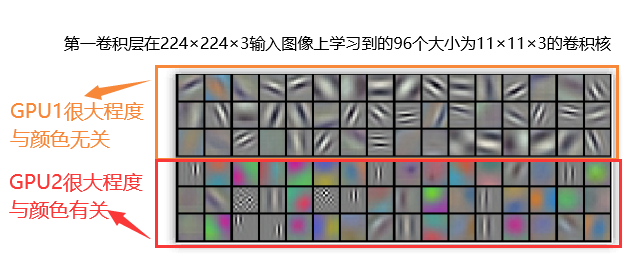
2．在机器视觉中，最底层可能检测边缘、颜色和纹理。

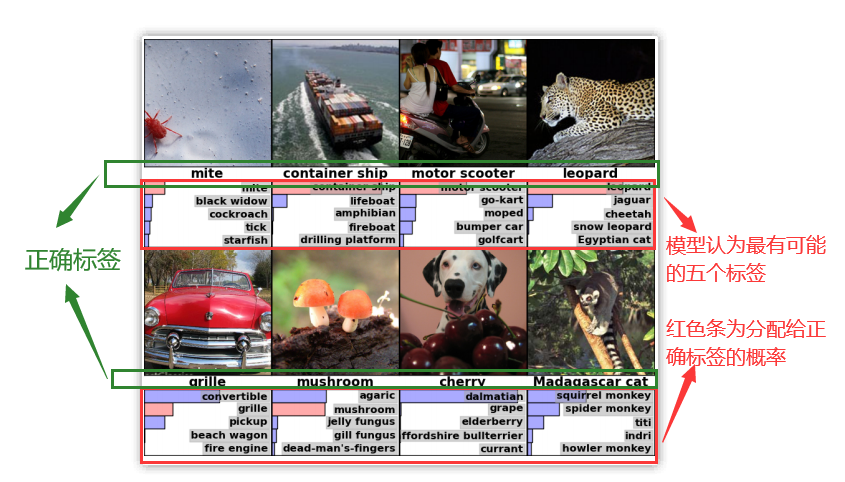
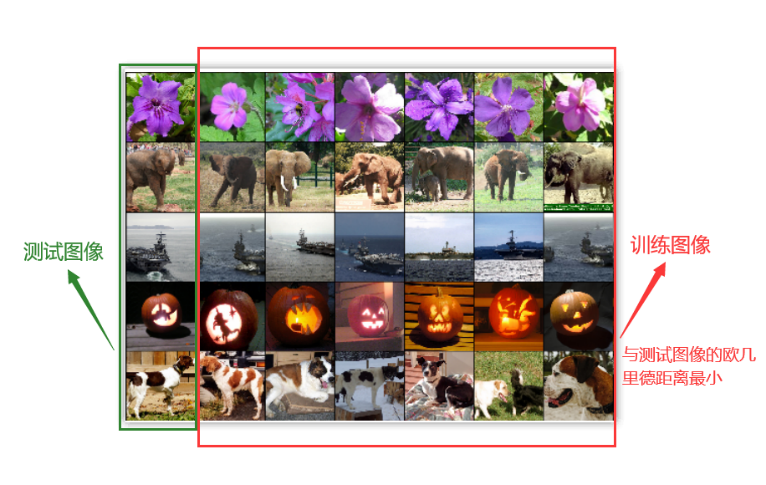
3．AlexNet的更高层建立在这些底层表示的基础上，以表示更大的特征，如眼睛、鼻子、草叶等等。

4．更高的层可以检测整个物体，如人、飞机、狗或飞盘。

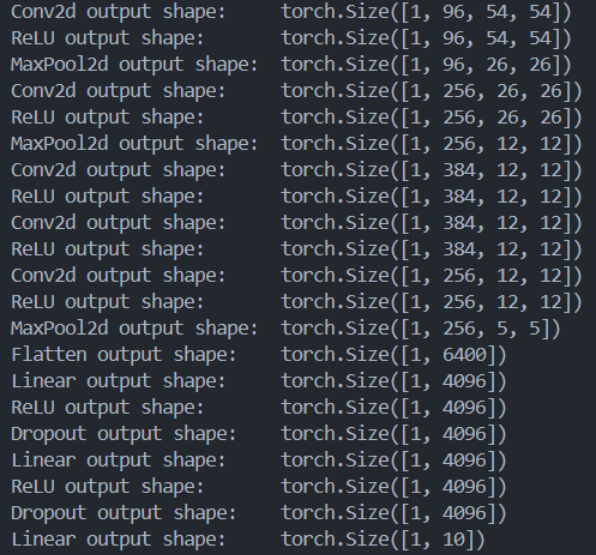
5．最终的隐藏神经元可以学习图像的综合表示，从而使属于不同类别的数据易于区分。

GPU 1 上的内核很大程度上与颜色无关，而 GPU 2 上的内核主要与颜色相关



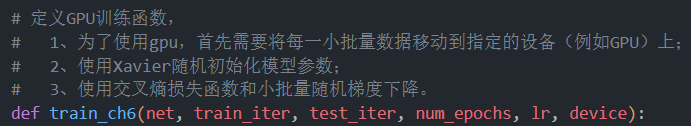
2.如果两个图像产生具有较小欧几里德分离的特征激活向量，我们可以说神经网络的更高层认为它们是相似的



七、代码讲解

下面与LeNet训练差不多，只是图片大小和模型架构不一样







这里使用的是Fashion-MNIST数据集以节省时间！！！

Fashion-MNIST图像的分辨率（28×28像素）低于ImageNet图像，将它们增加到224×224

通道数应该是3，但是Fashion-MNIST是1，ImageNet是3