一．AlexNe

AlexNet将LeNet的思想发扬光大，把CNN的基本原理应用到了很深很宽的网络中。AlexNet主要使用到的新技术点如下。

（1）成功使用ReLU作为CNN的激活函数，并验证其效果在较深的网络超过了Sigmoid，成功解决了Sigmoid在网络较深时的梯度弥散问题。虽然ReLU激活函数在很久之前就被提出了，但是直到AlexNet的出现才将其发扬光大。

（2）训练时使用Dropout随机忽略一部分神经元，以避免模型过拟合。Dropout虽有单独的论文论述，但是AlexNet将其实用化，通过实践证实了它的效果。在AlexNet中主要是最后几个全连接层使用了Dropout。

（3）在CNN中使用重叠的最大池化。此前CNN中普遍使用平均池化，AlexNet全部使用最大池化，避免平均池化的模糊化效果。并且AlexNet中提出让步长比池化核的尺寸小，这样池化层的输出之间会有重叠和覆盖，提升了特征的丰富性。

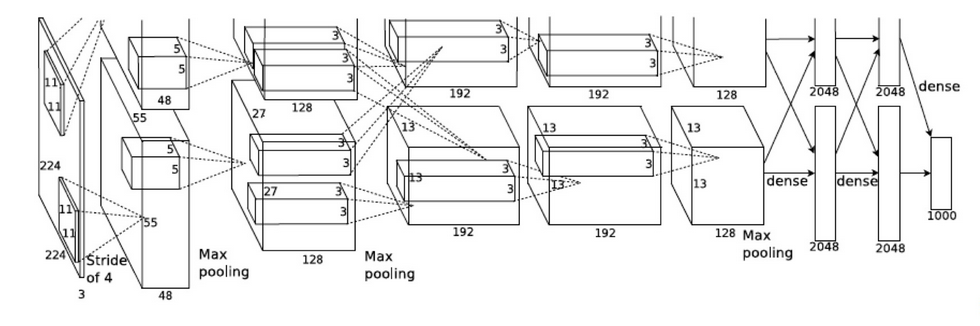
（4）提出了LRN层，对局部神经元的活动创建竞争机制，使得其中响应比较大的值变得相对更大，并抑制其他反馈较小的神经元，增强了模型的泛化能力。

（5）使用CUDA加速深度卷积网络的训练，利用GPU强大的并行计算能力，处理神经网络训练时大量的矩阵运算。AlexNet使用了两块GTX 580 GPU进行训练，单个GTX 580只有3GB显存，这限制了可训练的网络的最大规模。因此作者将AlexNet分布在两个GPU上，在每个GPU的显存中储存一半的神经元的参数。因为GPU之间通信方便，可以互相访问显存，而不需要通过主机内存，所以同时使用多块GPU也是非常高效的。同时，AlexNet的设计让GPU之间的通信只在网络的某些层进行，控制了通信的性能损耗。

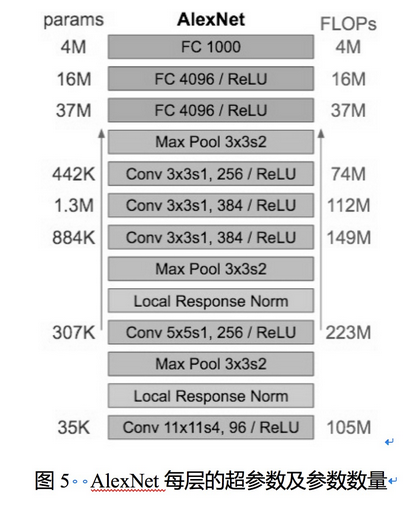
（6）数据增强，随机地从256\*256的原始图像中截取224\*224大小的区域（以及水平翻转的镜像），相当于增加了(256-224)\* (256-224)\*2=32\*32\*2=2048倍的数据量(这里乘2表示水平翻转增加一倍)。如果没有数据增强，仅靠原始的数据量，参数众多的CNN会陷入过拟合中，使用了数据增强后可以大大减轻过拟合，提升泛化能力。进行预测时，则是取图片的四个角加中间共5个位置，并进行左右翻转，一共获得10张图片，对他们进行预测并对10次结果求均值。同时，AlexNet论文中提到了会对图像的RGB数据进行PCA处理，并对主成分做一个标准差为0.1的高斯扰动，增加一些噪声，这个Trick可以让错误率再下降1%。

整个AlexNet有8个需要训练参数的层（不包括池化层和LRN层），前5层为卷积层，后3层为全连接层，如下图所示。AlexNet最后一层是有1000类输出的Softmax层用作分类。 LRN层出现在第1个及第2个卷积层后，而最大池化层出现在两个LRN层及最后一个卷积层后。ReLU激活函数则应用在这8层每一层的后面。因为AlexNet训练时使用了两块GPU，因此这个结构图中不少组件都被拆为了两部分。现在我们GPU的显存可以放下全部模型参数，因此只考虑一块GPU的情况即可。

AlexNet每层的超参数如图5所示。其中输入的图片尺寸为224\*224，第一个卷积层使用了较大的卷积核尺寸11\*11，步长为4，有96个卷积核；紧接着一个LRN层；然后是一个3\*3的最大池化层，步长为2。这之后的卷积核尺寸都比较小，都是5\*5或者3\*3的大小，并且步长都为1，即会扫描全图所有像素；而最大池化层依然保持为3\*3，并且步长为2。



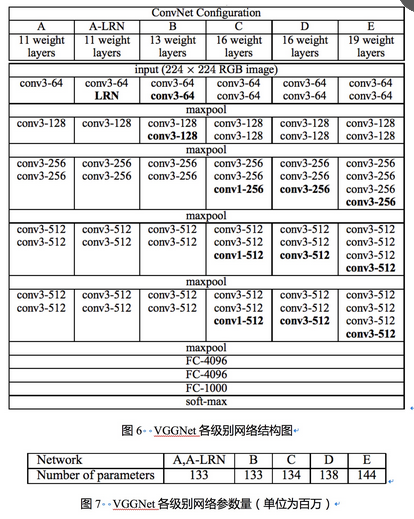
我们可以发现一个比较有意思的现象，在前几个卷积层，虽然计算量很大，但参数量很小，都在1M左右甚至更小，只占AlexNet总参数量的很小一部分。这就是卷积层有用的地方，可以通过较小的参数量提取有效的特征。而如果前几层直接使用全连接层，那么参数量和计算量将成为天文数字。虽然每一个卷积层占整个网络的参数量的1%都不到，但是如果去掉任何一个卷积层，都会使网络的分类性能大幅地下降。



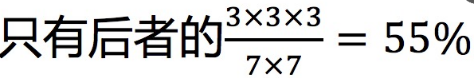
二．VGGNet

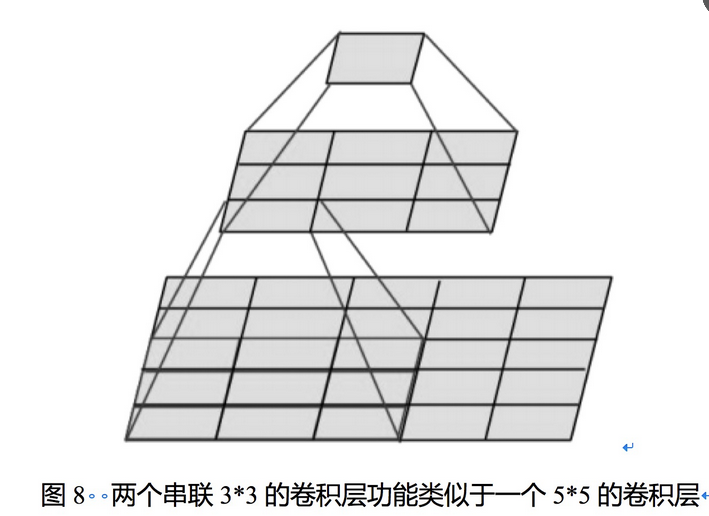
下面开始介绍ILSVRC2014年亚军 VGGNet。VGGNet是牛津大学计算机视觉组（Visual Geometry Group）和Google DeepMind公司的研究员一起研发的的深度卷积神经网络。VGGNet探索了卷积神经网络的深度与其性能之间的关系，通过反复堆叠3\*3的小型卷积核和2\*2的最大池化层，VGGNet成功地构筑了16~19层深的卷积神经网络。VGGNet相比之前state-of-the-art的网络结构，错误率大幅下降，并取得了ILSVRC 2014比赛分类项目的第2名和定位项目的第1名。同时VGGNet的拓展性很强，迁移到其他图片数据上的泛化性非常好。VGGNet的结构非常简洁，整个网络都使用了同样大小的卷积核尺寸（3\*3）和最大池化尺寸（2\*2）。到目前为止，VGGNet依然经常被用来提取图像特征。VGGNet训练后的模型参数在其官方网站上开源了，可用来在domain specific的图像分类任务上进行再训练（相当于提供了非常好的初始化权重），因此被用在了很多地方。

VGGNet论文中全部使用了3\*3的卷积核和2\*2的池化核，通过不断加深网络结构来提升性能。图6所示为VGGNet各级别的网络结构图，图7所示为每一级别的参数量，从11层的网络一直到19层的网络都有详尽的性能测试。虽然从A到E每一级网络逐渐变深，但是网络的参数量并没有增长很多，这是因为参数量主要都消耗在最后3个全连接层。前面的卷积部分虽然很深，但是消耗的参数量不大，不过训练比较耗时的部分依然是卷积，因其计算量比较大。这其中的D、E也就是我们常说的VGGNet-16和VGGNet-19。C很有意思，相比B多了几个1\*1的卷积层，1\*1卷积的意义主要在于线性变换，而输入通道数和输出通道数不变，没有发生降维。



VGGNet拥有5段卷积，每一段内有2~3个卷积层，同时每段尾部会连接一个最大池化层用来缩小图片尺寸。每段内的卷积核数量一样，越靠后的段的卷积核数量越多：64 – 128 – 256 – 512 – 512。其中经常出现多个完全一样的3\*3的卷积层堆叠在一起的情况，这其实是非常有用的设计。

如图8所示，两个3\*3的卷积层串联相当于1个5\*5的卷积层，即一个像素会跟周围5\*5的像素产生关联，可以说感受野大小为5\*5。而3个3\*3的卷积层串联的效果则相当于1个7\*7的卷积层。除此之外，3个串联的3\*3的卷积层，拥有比1个7\*7的卷积层更少的参数量，。最重要的是，3个3\*3的卷积层拥有比1个7\*7的卷积层更多的非线性变换（前者可以使用三次ReLU激活函数，而后者只有一次），使得CNN对特征的学习能力更强。



VGGNet在训练时有一个小技巧，先训练级别A的简单网络，再复用A网络的权重来初始化后面的几个复杂模型，这样训练收敛的速度更快。

在预测时，VGG采用Multi-Scale的方法，将图像scale到一个尺寸Q，并将图片输入卷积网络计算。然后在最后一个卷积层使用滑窗的方式进行分类预测，将不同窗口的分类结果平均，再将不同尺寸Q的结果平均得到最后结果，这样可提高图片数据的利用率并提升预测准确率。

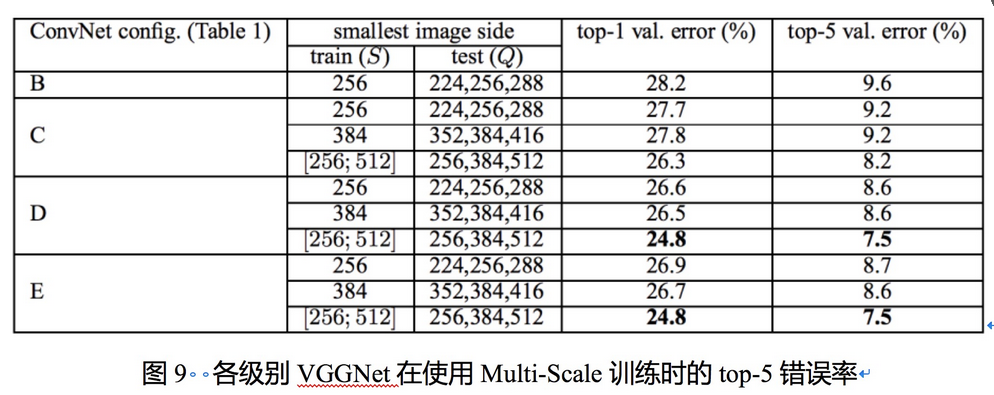
同时在训练中，VGGNet还使用了Multi-Scale的方法做数据增强，将原始图像缩放到不同尺寸S，然后再随机裁切224\*224的图片，这样能增加很多数据量，对于防止模型过拟合有很不错的效果。实践中，作者令S在[256,512]这个区间内取值，使用Multi-Scale获得多个版本的数据，并将多个版本的数据合在一起进行训练。

图9所示为VGGNet使用Multi-Scale训练时得到的结果，可以看到D和E都可以达到7.5%的错误率。最终提交到ILSVRC 2014的版本是仅使用Single-Scale的6个不同等级的网络与Multi-Scale的D网络的融合，达到了7.3%的错误率。不过比赛结束后作者发现只融合Multi-Scale的D和E可以达到更好的效果，错误率达到7.0%，再使用其他优化策略最终错误率可达到6.8%左右，非常接近同年的冠军Google Inceptin Net。同时，作者在对比各级网络时总结出了以下几个观点。

（1）LRN层作用不大。

（2）越深的网络效果越好。

（3）1\*1的卷积也是很有效的，但是没有3\*3的卷积好，大一些的卷积核可以学习更大的空间特征。



三．GoogleNet

下面介绍2014年ILSVRC比赛的冠军， Google Inception Net

Google Inception Net首次出现在ILSVRC 2014的比赛中（和VGGNet同年），就以较大优势取得了第一名。那届比赛中的Inception Net通常被称为Inception V1，它最大的特点是控制了计算量和参数量的同时，获得了非常好的分类性能——top-5错误率6.67%，只有AlexNet的一半不到。Inception V1有22层深，比AlexNet的8层或者VGGNet的19层还要更深。但其计算量只有15亿次浮点运算，同时只有500万的参数量，仅为AlexNet参数量（6000万）的1/12，却可以达到远胜于AlexNet的准确率，可以说是非常优秀并且非常实用的模型。

inception V1降低参数量的目的有两点，第一，参数越多模型越庞大，需要供模型学习的数据量就越大，而目前高质量的数据非常昂贵；第二，参数越多，耗费的计算资源也会更大。

Inception V1参数少但效果好的原因除了模型层数更深、表达能力更强外，还有两点：

一是去除了最后的全连接层，用全局平均池化层（即将图片尺寸变为1\*1）来取代它。全连接层几乎占据了AlexNet或VGGNet中90%的参数量，而且会引起过拟合，去除全连接层后模型训练更快并且减轻了过拟合。用全局平均池化层取代全连接层的做法借鉴了Network In Network（以下简称NIN）论文。

二是Inception V1中精心设计的Inception Module提高了参数的利用效率，其结构如图10所示。这一部分也借鉴了NIN的思想，形象的解释就是Inception Module本身如同大网络中的一个小网络，其结构可以反复堆叠在一起形成大网络。不过Inception V1比NIN更进一步的是增加了分支网络，NIN则主要是级联的卷积层和MLPConv层。一般来说卷积层要提升表达能力，主要依靠增加输出通道数，但副作用是计算量增大和过拟合。每一个输出通道对应一个滤波器，同一个滤波器共享参数，只能提取一类特征，因此一个输出通道只能做一种特征处理。而NIN中的MLPConv则拥有更强大的能力，允许在输出通道之间组合信息，因此效果明显。可以说，MLPConv基本等效于普通卷积层后再连接1\*1的卷积和ReLU激活函数。

我们再来看Inception Module的基本结构，其中有4个分支：

第一个分支对输入进行1\*1的卷积，这其实也是NIN中提出的一个重要结构。1\*1的卷积是一个非常优秀的结构，它可以跨通道组织信息，提高网络的表达能力，同时可以对输出通道升维和降维。可以看到Inception Module的4个分支都用到了1\*1卷积，来进行低成本（计算量比3\*3小很多）的跨通道的特征变换。

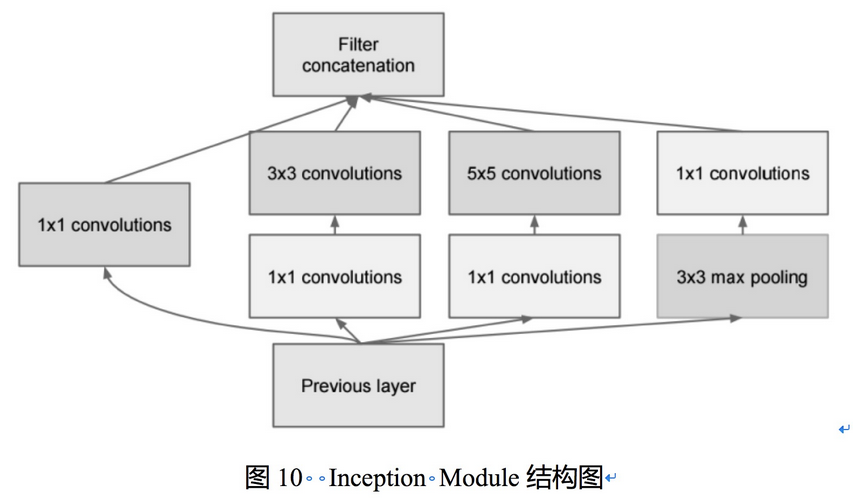
第二个分支先使用了1\*1卷积，然后连接3\*3卷积，相当于进行了两次特征变换。

第三个分支类似，先是1\*1的卷积，然后连接5\*5卷积。

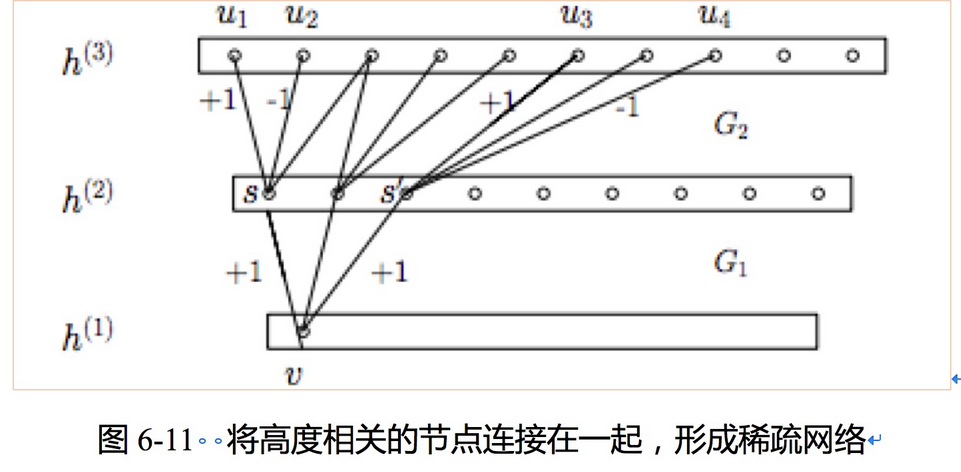
最后一个分支则是3\*3最大池化后直接使用1\*1卷积。

我们可以发现，有的分支只使用1\*1卷积，有的分支使用了其他尺寸的卷积时也会再使用1\*1卷积，这是因为1\*1卷积的性价比很高，用很小的计算量就能增加一层特征变换和非线性化。

Inception Module的4个分支在最后通过一个聚合操作合并（在输出通道数这个维度上聚合）。Inception Module中包含了3种不同尺寸的卷积和1个最大池化，增加了网络对不同尺度的适应性，这一部分和Multi-Scale的思想类似。早期计算机视觉的研究中，受灵长类神经视觉系统的启发，Serre使用不同尺寸的Gabor滤波器处理不同尺寸的图片，Inception V1借鉴了这种思想。Inception V1的论文中指出，Inception Module可以让网络的深度和宽度高效率地扩充，提升准确率且不致于过拟合。



人脑神经元的连接是稀疏的，因此研究者认为大型神经网络的合理的连接方式应该也是稀疏的。稀疏结构是非常适合神经网络的一种结构，尤其是对非常大型、非常深的神经网络，可以减轻过拟合并降低计算量，例如卷积神经网络就是稀疏的连接。Inception Net的主要目标就是找到最优的稀疏结构单元（即Inception Module），论文中提到其稀疏结构基于Hebbian原理，这里简单解释一下Hebbian原理：神经反射活动的持续与重复会导致神经元连接稳定性的持久提升，当两个神经元细胞A和B距离很近，并且A参与了对B重复、持续的兴奋，那么某些代谢变化会导致A将作为能使B兴奋的细胞。总结一下即“一起发射的神经元会连在一起”（Cells that fire together, wire together），学习过程中的刺激会使神经元间的突触强度增加。受Hebbian原理启发，另一篇文章Provable Bounds for Learning Some Deep Representations提出，如果数据集的概率分布可以被一个很大很稀疏的神经网络所表达，那么构筑这个网络的最佳方法是逐层构筑网络：将上一层高度相关（correlated）的节点聚类，并将聚类出来的每一个小簇（cluster）连接到一起，如图11所示。这个相关性高的节点应该被连接在一起的结论，即是从神经网络的角度对Hebbian原理有效性的证明。



因此一个“好”的稀疏结构，应该是符合Hebbian原理的，我们应该把相关性高的一簇神经元节点连接在一起。在普通的数据集中，这可能需要对神经元节点聚类，但是在图片数据中，天然的就是临近区域的数据相关性高，因此相邻的像素点被卷积操作连接在一起。而我们可能有多个卷积核，在同一空间位置但在不同通道的卷积核的输出结果相关性极高。因此，一个1\*1的卷积就可以很自然地把这些相关性很高的、在同一个空间位置但是不同通道的特征连接在一起，这就是为什么1\*1卷积这么频繁地被应用到Inception Net中的原因。

1\*1卷积所连接的节点的相关性是最高的，而稍微大一点尺寸的卷积，比如3\*3、5\*5的卷积所连接的节点相关性也很高，因此也可以适当地使用一些大尺寸的卷积，增加多样性（diversity）。最后Inception Module通过4个分支中不同尺寸的1\*1、3\*3、5\*5等小型卷积将相关性很高的节点连接在一起，就完成了其设计初衷，构建出了很高效的符合Hebbian原理的稀疏结构。

在Inception Module中，通常1\*1卷积的比例（输出通道数占比）最高，3\*3卷积和5\*5卷积稍低。而在整个网络中，会有多个堆叠的Inception Module，我们希望靠后的Inception Module可以捕捉更高阶的抽象特征，因此靠后的Inception Module的卷积的空间集中度应该逐渐降低，这样可以捕获更大面积的特征。因此，越靠后的Inception Module中，3\*3和5\*5这两个大面积的卷积核的占比（输出通道数）应该更多。

Inception Net有22层深，除了最后一层的输出，其中间节点的分类效果也很好。因此在Inception Net中，还使用到了辅助分类节点（auxiliary classifiers），即将中间某一层的输出用作分类，并按一个较小的权重（0.3）加到最终分类结果中。这样相当于做了模型融合，同时给网络增加了反向传播的梯度信号，也提供了额外的正则化，对于整个Inception Net的训练很有裨益。

当年的Inception V1还是跑在TensorFlow的前辈DistBelief上的，并且只运行在CPU上。当时使用了异步的SGD训练，学习速率每迭代8个epoch降低4%。同时，Inception V1也使用了Multi-Scale、Multi-Crop等数据增强方法，并在不同的采样数据上训练了7个模型进行融合，得到了最后的ILSVRC 2014的比赛成绩——top-5错误率6.67%。

同时，Google Inception Net还是一个大家族，包括：

2014年9月的论文Going Deeper with Convolutions提出的Inception V1（top-5错误率6.67%）。

2015年2月的论文Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate提出的Inception V2（top-5错误率4.8%）。

2015年12月的论文Rethinking the Inception Architecture for Computer Vision提出的Inception V3（top-5错误率3.5%）。

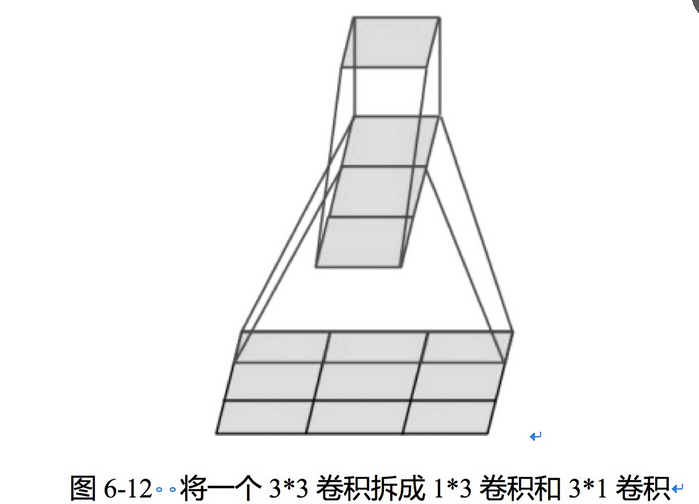
2016年2月的论文Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning提出的Inception V4（top-5错误率3.08%）。

Inception V2学习了VGGNet，用两个3\*3的卷积代替5\*5的大卷积（用以降低参数量并减轻过拟合），还提出了著名的Batch Normalization（以下简称BN）方法。BN是一个非常有效的正则化方法，可以让大型卷积网络的训练速度加快很多倍，同时收敛后的分类准确率也可以得到大幅提高。BN在用于神经网络某层时，会对每一个mini-batch数据的内部进行标准化（normalization）处理，使输出规范化到N(0,1)的正态分布，减少了Internal Covariate Shift（内部神经元分布的改变）。

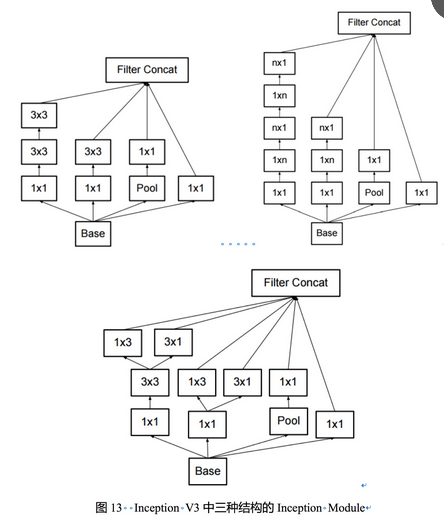
BN的论文指出，传统的深度神经网络在训练时，每一层的输入的分布都在变化，导致训练变得困难，我们只能使用一个很小的学习速率解决这个问题。而对每一层使用BN之后，我们就可以有效地解决这个问题，学习速率可以增大很多倍，达到之前的准确率所需要的迭代次数只有1/14，训练时间大大缩短。而达到之前的准确率后，可以继续训练，并最终取得远超于Inception V1模型的性能——top-5错误率4.8%，已经优于人眼水平。因为BN某种意义上还起到了正则化的作用，所以可以减少或者取消Dropout，简化网络结构。

当然，只是单纯地使用BN获得的增益还不明显，还需要一些相应的调整：增大学习速率并加快学习衰减速度以适用BN规范化后的数据；去除Dropout并减轻L2正则（因BN已起到正则化的作用）；去除LRN；更彻底地对训练样本进行shuffle；减少数据增强过程中对数据的光学畸变（因为BN训练更快，每个样本被训练的次数更少，因此更真实的样本对训练更有帮助）。在使用了这些措施后，Inception V2在训练达到Inception V1的准确率时快了14倍，并且模型在收敛时的准确率上限更高。

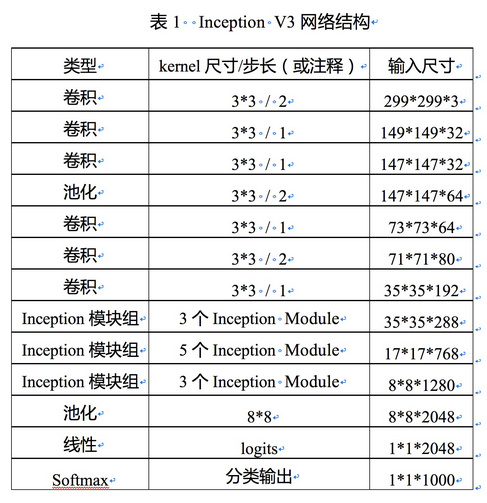
Inception V3网络则主要有两方面的改造：一是引入了Factorization into small convolutions的思想，将一个较大的二维卷积拆成两个较小的一维卷积，比如将7\*7卷积拆成1\*7卷积和7\*1卷积，或者将3\*3卷积拆成1\*3卷积和3\*1卷积，如图12所示。一方面节约了大量参数，加速运算并减轻了过拟合（比将7\*7卷积拆成1\*7卷积和7\*1卷积，比拆成3个3\*3卷积更节约参数），同时增加了一层非线性扩展模型表达能力。论文中指出，这种非对称的卷积结构拆分，其结果比对称地拆为几个相同的小卷积核效果更明显，可以处理更多、更丰富的空间特征，增加特征多样性。



另一方面，Inception V3优化了Inception Module的结构，现在Inception Module有35\*35、17\*17和8\*8三种不同结构，如图13所示。这些Inception Module只在网络的后部出现，前部还是普通的卷积层。并且Inception V3除了在Inception Module中使用分支，还在分支中使用了分支（8\*8的结构中），可以说是Network In Network In Network。



Inception V3其整个网络结构如表1所示。



Inception V4相比V3主要是结合了微软的ResNet，而ResNet将在下一节单独讲解，这里不多做赘述。

四．ResNet

下面介绍ILSVRC 2015年冠军，微软的ResNet

ResNet（Residual Neural Network）由微软研究院的Kaiming He等4名华人提出，通过使用Residual Unit成功训练152层深的神经网络，在ILSVRC 2015比赛中获得了冠军，取得3.57%的top-5错误率，同时参数量却比VGGNet低，效果非常突出。ResNet的结构可以极快地加速超深神经网络的训练，模型的准确率也有非常大的提升。上一节我们讲解了Inception V3，而Inception V4则是将Inception Module和ResNet相结合。可以看到ResNet是一个推广性非常好的网络结构，甚至可以直接应用到Inception Net中。

在ResNet之前，瑞士教授Schmidhuber提出了Highway Network，原理与ResNet很相似。这位Schmidhuber教授同时也是LSTM网络的发明者，而且是早在1997年发明的，可谓是神经网络领域元老级的学者。通常认为神经网络的深度对其性能非常重要，但是网络越深其训练难度越大，Highway Network的目标就是解决极深的神经网络难以训练的问题。

Highway Network相当于修改了每一层的激活函数，此前的激活函数只是对输入做一个非线性变换，Highway NetWork则允许保留一定比例的原始输入x，即，其中T为变换系数，C为保留系数。这样前面一层的信息，有一定比例可以不经过矩阵乘法和非线性变换，直接传输到下一层，仿佛一条信息高速公路，因此得名Highway Network。Highway Network主要通过gating units学习如何控制网络中的信息流，即学习原始信息应保留的比例。这个可学习的gating机制，正是借鉴自Schmidhuber教授早年的LSTM循环神经网络中的gating。几百乃至上千层深的Highway Network可以直接使用梯度下降算法训练，并可以配合多种非线性激活函数，学习极深的神经网络现在变得可行了。事实上，Highway Network的设计在理论上允许其训练任意深的网络，其优化方法基本上与网络的深度独立，而传统的神经网络结构则对深度非常敏感，训练复杂度随深度增加而急剧增加。

ResNet和HighWay Network非常类似，也是允许原始输入信息直接传输到后面的层中。ResNet最初的灵感出自这个问题：在不断加神经网络的深度时，会出现一个Degradation的问题，即准确率会先上升然后达到饱和，再持续增加深度则会导致准确率下降。这并不是过拟合的问题，因为不光在测试集上误差增大，训练集本身误差也会增大。假设有一个比较浅的网络达到了饱和的准确率，那么后面再加上几个的全等映射层，起码误差不会增加，即更深的网络不应该带来训练集上误差上升。而这里提到的使用全等映射直接将前一层输出传到后面的思想，就是ResNet的灵感来源。假定某段神经网络的输入是x，期望输出是，如果我们直接把输入x传到输出作为初始结果，那么此时我们需要学习的目标就是F(x)。如图14所示，这就是一个ResNet的残差学习单元（Residual Unit），ResNet相当于将学习目标改变了，不再是学习一个完整的输出，只是输出和输入的差别，即残差。

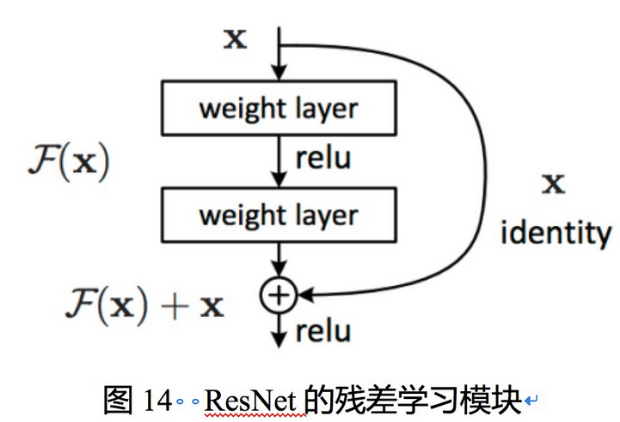
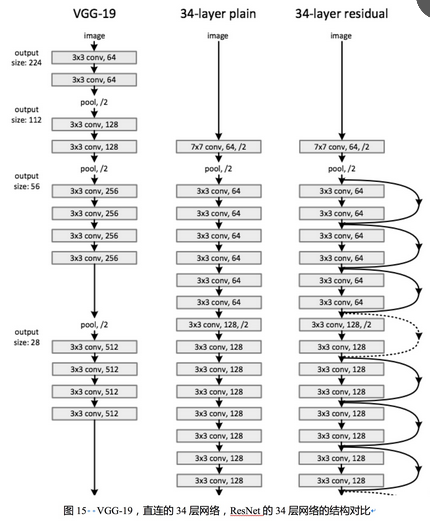


图15所示为VGGNet-19，以及一个34层深的普通卷积网络，和34层深的ResNet网络的对比图。可以看到普通直连的卷积神经网络和ResNet的最大区别在于，ResNet有很多旁路的支线将输入直接连到后面的层，使得后面的层可以直接学习残差，这种结构也被称为shortcut或skip connections。



传统的卷积层或全连接层在信息传递时，或多或少会存在信息丢失、损耗等问题。ResNet在某种程度上解决了这个问题，通过直接将输入信息绕道传到输出，保护信息的完整性，整个网络则只需要学习输入、输出差别的那一部分，简化学习目标和难度。

在ResNet的论文中，除了提出图14中的两层残差学习单元，还有三层的残差学习单元。两层的残差学习单元中包含两个相同输出通道数（因为残差等于目标输出减去输入，即，因此输入、输出维度需保持一致）的3\*3卷积；而3层的残差网络则使用了Network In Network和Inception Net中的1\*1卷积，并且是在中间3\*3的卷积前后都使用了1\*1卷积，有先降维再升维的操作。另外，如果有输入、输出维度不同的情况，我们可以对x做一个线性映射变换维度，再连接到后面的层。

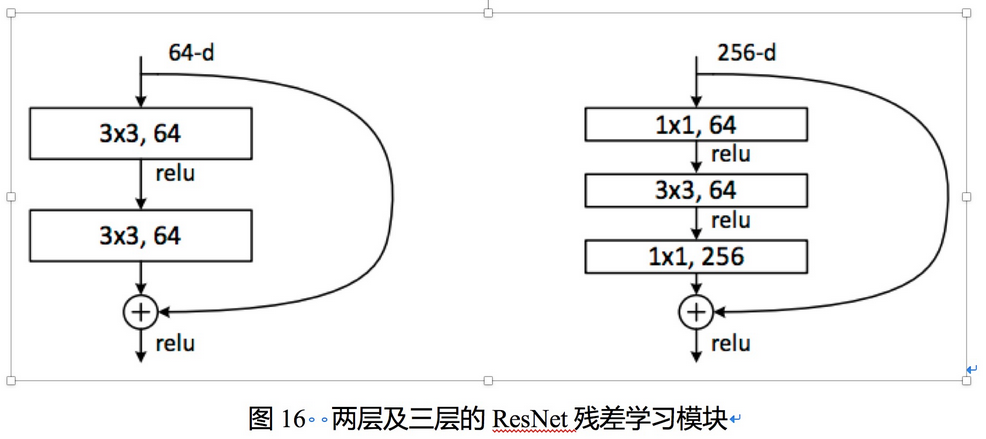
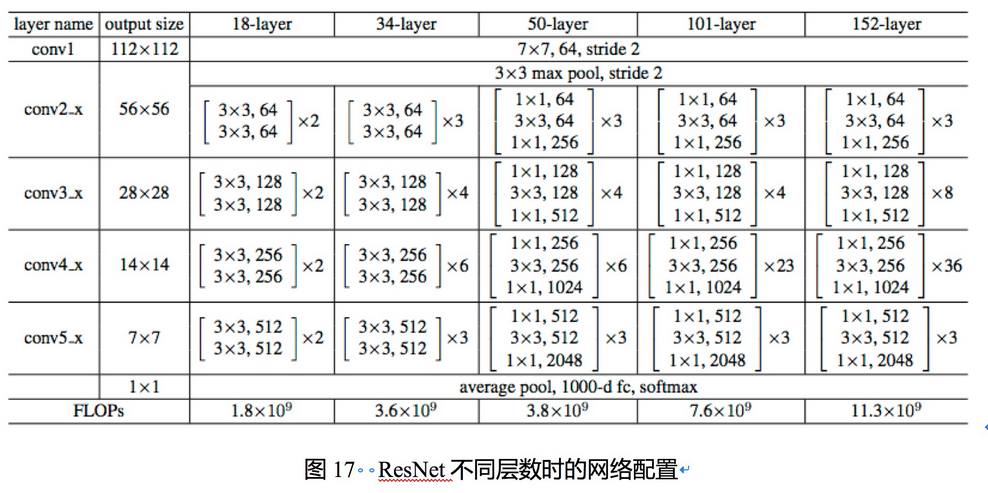


图17所示为ResNet在不同层数时的网络配置，其中基础结构很类似，都是前面提到的两层和三层的残差学习单元的堆叠。



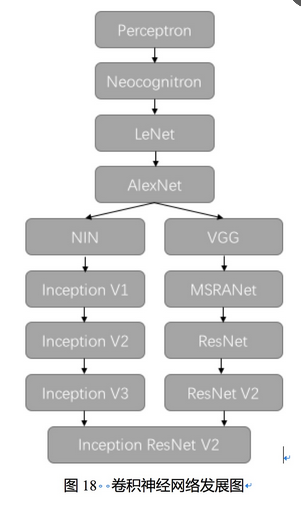
在使用了ResNet的结构后，可以发现层数不断加深导致的训练集上误差增大的现象被消除了，ResNet网络的训练误差会随着层数增大而逐渐减小，并且在测试集上的表现也会变好。在ResNet推出后不久，Google就借鉴了ResNet的精髓，提出了Inception V4和Inception-ResNet-V2，并通过融合这两个模型，在ILSVRC数据集上取得了惊人的3.08%的错误率。可见，ResNet及其思想对卷积神经网络研究的贡献确实非常显著，具有很强的推广性。

在ResNet的作者的第二篇相关论文Identity Mappings in Deep Residual Networks中，ResNet V2被提出。ResNet V2和ResNet V1的主要区别在于，作者通过研究ResNet残差学习单元的传播公式，发现前馈和反馈信号可以直接传输，因此skip connection的非线性激活函数（如ReLU）替换为Identity Mappings（）。同时，ResNet V2在每一层中都使用了Batch Normalization。这样处理之后，新的残差学习单元将比以前更容易训练且泛化性更强。

根据Schmidhuber教授的观点，ResNet类似于一个没有gates的LSTM网络，即将输入x传递到后面层的过程是一直发生的，而不是学习出来的。同时，最近也有两篇论文表示，ResNet基本等价于RNN且ResNet的效果类似于在多层网络间的集成方法（ensemble）。ResNet在加深网络层数上做出了重大贡献，而另一篇论文The Power of Depth for Feedforward Neural Networks则从理论上证明了加深网络比加宽网络更有效，算是给ResNet提供了声援，也是给深度学习为什么要深才有效提供了合理解释。

五．总结

到此就将ALexNet、VGGNet、Inception Net、ResNet四种经典的卷积介绍完了，下面我们简单总价一下。



我们简单回顾卷积神经网络的历史，图6-18所示大致勾勒出最近几十年卷积神经网络的发展方向。Perceptron（感知机）于1957年由Frank Resenblatt提出，而Perceptron不仅是卷积网络，也是神经网络的始祖。Neocognitron（神经认知机）是一种多层级的神经网络，由日本科学家Kunihiko Fukushima于20世纪80年代提出，具有一定程度的视觉认知的功能，并直接启发了后来的卷积神经网络。LeNet-5由CNN之父Yann LeCun于1997年提出，首次提出了多层级联的卷积结构，可对手写数字进行有效识别。

可以看到前面这三次关于卷积神经网络的技术突破，间隔时间非常长，需要十余年甚至更久才出现一次理论创新。而后于2012年，Hinton的学生Alex依靠8层深的卷积神经网络一举获得了ILSVRC 2012比赛的冠军，瞬间点燃了卷积神经网络研究的热潮。AlexNet成功应用了ReLU激活函数、Dropout、最大覆盖池化、LRN层、GPU加速等新技术，并启发了后续更多的技术创新，卷积神经网络的研究从此进入快车道。

在AlexNet之后，我们可以将卷积神经网络的发展分为两类，一类是网络结构上的改进调整（图6-18中的左侧分支），另一类是网络深度的增加（图18中的右侧分支）。 2013年，颜水成教授的Network in Network工作首次发表，优化了卷积神经网络的结构，并推广了1\*1的卷积结构。在改进卷积网络结构的工作中，后继者还有2014年的Google Inception Net V1，提出了Inception Module这个可以反复堆叠的高效的卷积网络结构，并获得了当年ILSVRC比赛的冠军。

2015年初的Inception V2提出了Batch Normalization，大大加速了训练过程，并提升了网络性能。2015年年末的Inception V3则继续优化了网络结构，提出了Factorization in Small Convolutions的思想，分解大尺寸卷积为多个小卷积乃至一维卷积。而另一条分支上，许多研究工作则致力于加深网络层数，2014年，ILSVRC比赛的亚军VGGNet全程使用3\*3的卷积，成功训练了深达19层的网络，当年的季军MSRA-Net也使用了非常深的网络。

2015年，微软的ResNet成功训练了152层深的网络，一举拿下了当年ILSVRC比赛的冠军，top-5错误率降低至3.46%。其后又更新了ResNet V2，增加了Batch Normalization，并去除了激活层而使用Identity Mapping或Preactivation，进一步提升了网络性能。此后，Inception ResNet V2融合了Inception Net优良的网络结构，和ResNet训练极深网络的残差学习模块，集两个方向之长，取得了更好的分类效果。

我们可以看到，自AlexNet于2012年提出后，深度学习领域的研究发展极其迅速，基本上每年甚至每几个月都会出现新一代的技术。新的技术往往伴随着新的网络结构，更深的网络的训练方法等，并在图像识别等领域不断创造新的准确率记录。至今，ILSVRC比赛和卷积神经网络的研究依然处于高速发展期，CNN的技术日新月异。