《Deep learning》 在线书籍：

<http://www.deeplearningbook.org/>

[**Neural Networks and Deep Learning**](http://neuralnetworksanddeeplearning.com/index.html)

Bengio 的《Deep learning》全书pdf版：

[deep learning.pdf](deep%20learning.pdf)

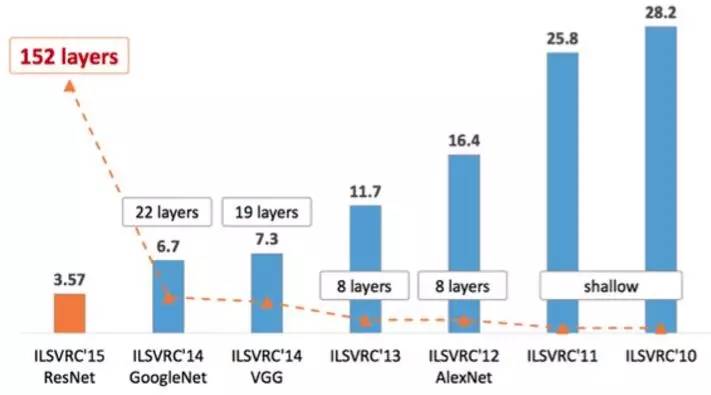
Deep learning 课程：

<https://github.com/Kulbear/deep-learning-coursera>

# ImageNet

[(http://www.infoq.com/cn/articles/cnn-and-imagenet-champion-model-analysis)]((http:/www.infoq.com/cn/articles/cnn-and-imagenet-champion-model-analysis))

ImageNet是斯坦福大学李飞飞教授主持设立的关于计算机视觉的数据库，里面含有大量的图片，这些图片分为上万个类别，是深度学习领域一个非常火热的竞赛。近年来由于硬件水平的提高，以及大数据量的爆发式增长，给深度学习提供了便利的发展条件，ImageNet的水平在近5年取得了飞速发展。特别是卷积神经网络（CNN）在计算机视觉领域用于图像分类和识别，在自然语言处理领域的语音识别，都取得了不错的效果。

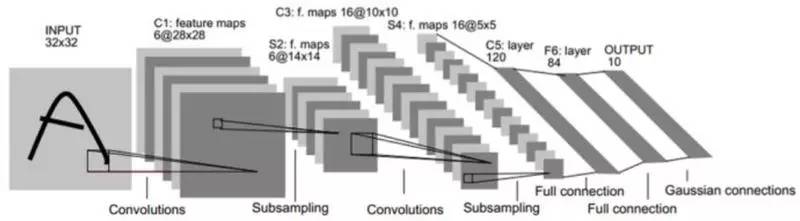


如图所示，ILSVRC的top-5错误率在最近几年取得重大突破，而主要的突破点都是在深度学习和卷积神经网络，成绩的大幅提升几乎都伴随着卷积神经网络的层数加深。

## [经典的卷积网络LeNet5](med-DL/Lecun98.pdf)

LeNet5当时的特性有如下几点。

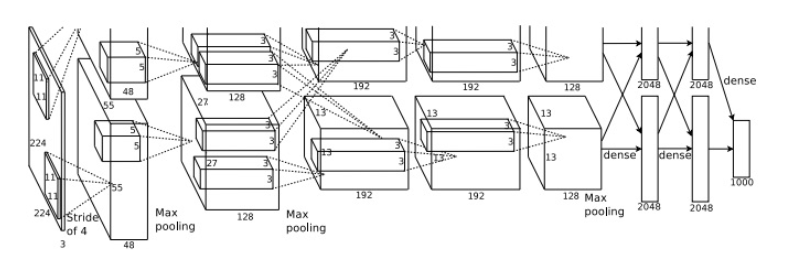
* 每个卷积层包含三个部分：卷积、池化和非线性激活函数
* 使用卷积提取空间特征
* 降采样（Subsample）的平均池化层（Average Pooling）
* 双曲正切（Tanh）或S型（Sigmoid）的激活函数
* MLP作为最后的分类器
* 层与层之间的稀疏连接减少计算复杂度



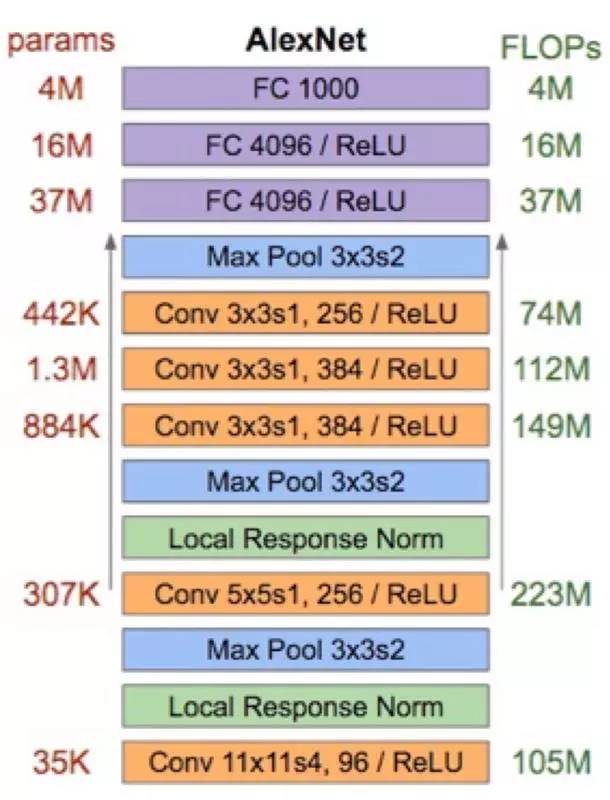
## [AlexNet（2012）](med-DL/imagenet.pdf)

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GTX580GPU进行训练，单个GTX580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。
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(256224)2\*2=2048倍的数据量。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。

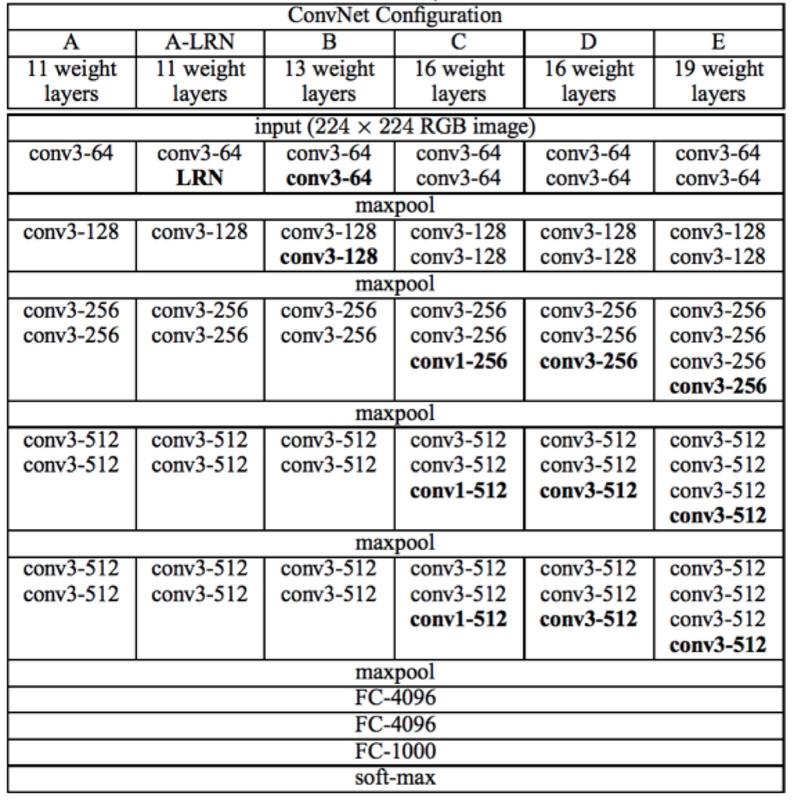


整个AlexNet有8个需要训练参数的层（不包括池化层和LRN层），前5层为卷积层，后3层为全连接层，上图所示。AlexNet最后一层是有1000类输出的Softmax层用作分类。LRN层出现在第1个及第2个卷积层后，而最大池化层出现在两个LRN层及最后一个卷积层后。



## [VGGNet（2014）](med-DL/1409.1556.pdf)

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。

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。下图所示为VGGNet各级别的网络结构图，和每一级别的参数量，从11层的网络一直到19层的网络都有详尽的性能测试。

作者在对比各级网络时总结出了以下几个观点。

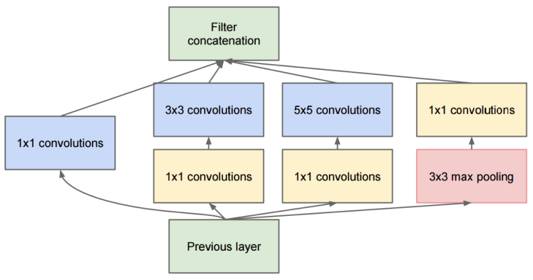
1. LRN层作用不大。
2. 越深的网络效果越好。
3. 1\*1的卷积也是很有效的，但是没有3\*3的卷积好，大一些的卷积核可以学习更大的空间特征。

## [InceptionNet（2014）](med-DL/1409.4842.pdf)

Google InceptionNet首次出现在ILSVRC2014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InceptionNet通常被称为InceptionV1，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。

InceptionV1参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。

二是InceptionV1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Network In Network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。



## [ResNet](med-DL/1512.03385.pdf)

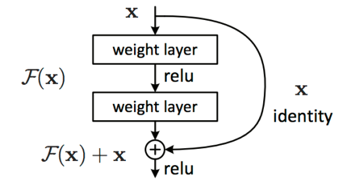
ResNet（Residual Neural Network）由微软研究院的Kaiming He等4名华人提出，通过使用Residual Unit成功训练152层深的神经网络，在ILSVRC2015比赛中获得了冠军，取得3.57%的top-5错误率，同时参数量却比VGGNet低，效果非常突出。

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。

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H(x)，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)=H(x)-x。



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H(x)，只是输出和输入的差别H(x)-x，即残差。

## 2017:

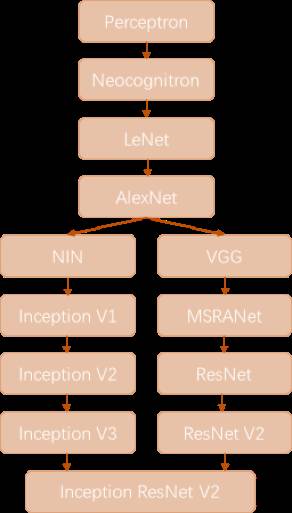
Squeeze-and-Excitation Networks

GitHub: <https://github.com/hujie-frank/SENet.git>

Paper: <https://link.zhihu.com/?target=https%3A//arxiv.org/abs/1709.01507>

(or local: [med-DL\1709.01507.pdf](med-DL/1709.01507.pdf))

# CNN 历史



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ILSVRC2012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（上图中的左侧分支），另一类是网络深度的增加（上图中的右侧分支）。

2013年，颜水成教授的[Network in Network](med-DL/1312.4400.pdf)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](med-DL/1502.03167.pdf)，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。

而右侧分支上，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%。

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。