**CNN发展史**

# 一、简述

最近一直在看深度学习相关的东西，而CNN是深度学习中的核心算法之一，也是2012年以来将人工智能推向风口浪尖的推手。今天我来与大家一起分享一下CNN的发展历史，看看各路大牛是如何解决问题，并造就了现在CNN网络机构百花齐放的状态。

1、卷积神经网络

首先来简单说一下，什么是CNN，卷积神经网络，它是一种人工神经网络的结构，是从猫的视觉神经结构中得到了灵感，进而模拟其结构设计出来的一种人工神经网络结构。（要了解CNN，需要先了解什么是人工神经网络算法，可以参考我的另一篇文章：[《【神经网络算法入门】详细推导全连接神经网络算法及反向传播算法+Python实现代码》](https://blog.csdn.net/cuiyuan605/article/details/84307323)）

这种结构通过卷积核来获取“感受野”范围内数据之间的关系特征。一张图片里，相邻的像素显然是有更强的相关性，相比于全连接神经网络，CNN突出了这种相邻的关系特征，因而更加准确的获取了图片内的有用信息。

具体卷积核是怎么工作的我找了几张图，大家感受一下：

[CNN1](CNN1.jpg)

[CNN2](CNN2.gif)

[CNN3](CNN3.gif)

[CNN4](CNN4.gif)

所以说，CNN的原理就是通过卷积核获取图像的空间关系特征，使用BP（反向传播）算法调整卷积核的参数，最终得到一个可以有效获取图片内有用信息的模型。

下面我们就从CNN的发展历史角度，来了解CNN是如何一步一步发展，在这过程中解决了什么问题。

2、比赛

讲历史就必须有一个主线，而我要说的这个历史主线就是ILSVRC，一个想了解CNN必须要了解的计算机视觉领域最权威的学术竞赛。

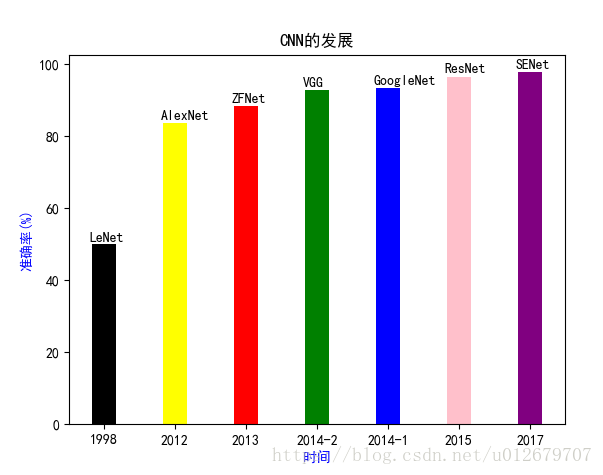
ILSVRC（ImageNet Large Scale Visual Recognition Challenge）竞赛(2010-2017)，这个竞赛所使用的数据集，是由斯坦福大学李飞飞教授主导，包含了超过1400万张全尺寸的有标记图片。ILSVRC比赛会每年从ImageNet数据集中抽出部分样本用于比赛，以2012年为例，比赛的训练集包含1281167张图片，验证集包含50000张图片，测试集为100000张图片。竞赛的项目主要包括图像的分类和定位（CLS-LOC）、目标检测（DET）、视频目标检测（VID）和场景分类（Scene）。

我下面讲的，主要是在最基本的分类问题上，历届表现优异的模型。不过ILSVRC从2017之后就停办了，另外一个计算机视觉领域的赛事，就成为了该领域权威的竞赛，它就是从2014年开始由微软举办的MS COCO竞赛。

MS COCO（Microsoft COCO: Common Objects in Context）竞赛（2014-） 物体检测、全景分割、人体关键点、姿态估计（DensePose）；2018年开始又新增了，街景检测、街景全景分割。在这里就不详细介绍了。

# 二、CNN发展史

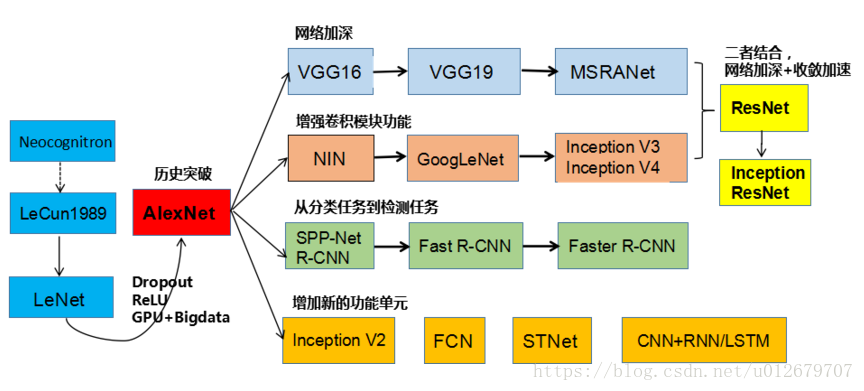
先偷一张图给大家看看，就是一系列里程碑式的模型，分别在ImageNet数据集图像分类的准确率表现：



这里面除了LeNet以外，其他都是在ILSVRC竞赛里表现优异的模型，为什么把LeNet放在这里面呢，我想来作者的想法一是做个对比，另外一个就是对LeNet表示一下尊重，因为LeNet是其他所有CNN网络模型的鼻祖，它的作者就是CNN之父、2018年图灵奖得主、人工智能三主神之一的Yann Lecun扬·莱坎，就是这个大叔：



下面我们来捋一捋各个模型之间的关系，于是我又偷了一张图：



2012年AlexNet的发布使人工智能算法迎来了历史性的突破，才造就了之后的舆论关注和CNN的研究爆发式的增长。之后一系列著名的CNN模型都是在AlexNet的基础上演变出来的，它门主要分为两个流派，一个是以加深网络为主的VGG流派，另一个是以增强卷积模块功能为主的NIN流派。之后两个流派相结合产生了Inception ResNet。

如果说AlexNet是之后所有CNN模型的爸爸，那么LeNet就是所有这些模型的爷爷。下面我们就从LeNet开始一一介绍各个经典模型都做了哪些工作，解决了什么问题。

下面的介绍中会提到一些经典的论文，这些论文可以从这里下载：

链接：https://pan.baidu.com/s/1e\_e3Lv0Y1WFPe88f9nv7nA

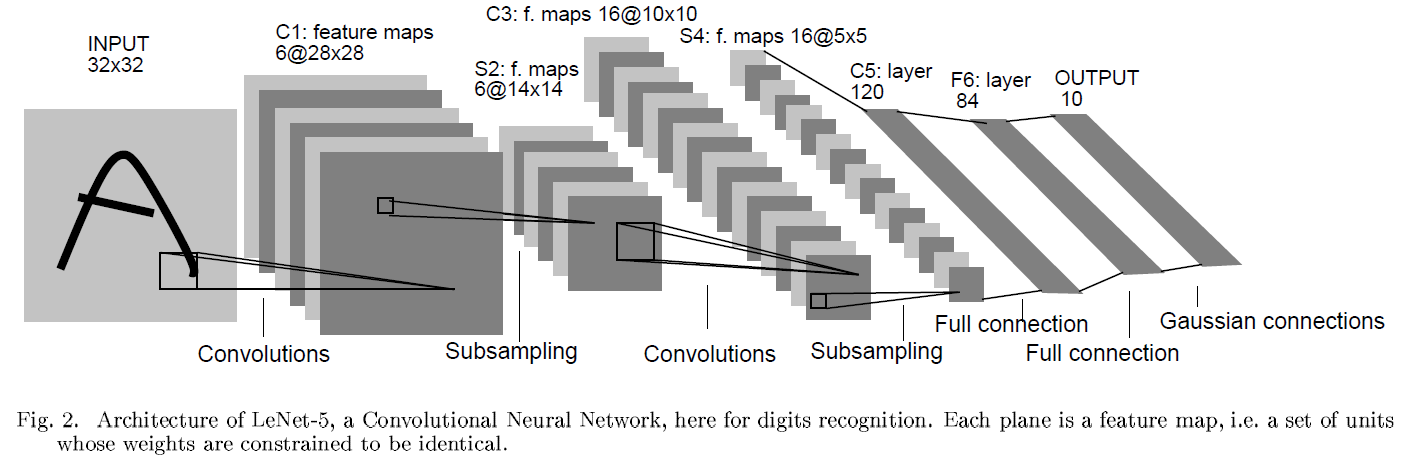
提取码：ha8e

供大家参考

1、LeNet（1998）--开山鼻祖

首先是LetNet，它是由CNN之父Lecun在1998年提出，用于解决手写数字识别任务的，著名的Tensorflow入门学习数据集MNIST，就是那时候做出来的。LeNet对应的论文是《Gradient-Based Learning Applied to Document Recognition》。

LeNet又叫LeNet-5，这个5是指它的网络结构中有5个表示层，具体结构如下图所示：



LeNet5包含Input、卷积层1、池化层1、卷积层2、池化层2、全连接层、输出层。

1998年的LeNet5模型的发布标注着CNN的真正面世，但是这个模型在后来的一段时间并未能火起来，主要原因是当时的计算力跟不上，而且其他的算法（比如SVM）也能达到类似的效果甚至超过。不过LeNet最大的贡献是：**定义了CNN卷积层、池化层、全连接层的基本结构，是CNN的鼻祖**。

 LeNet5当时的特性有如下几点：

（1）每个卷积层包含三个部分：卷积、池化和非线性激活函数

（2）使用卷积提取空间特征（起初被称为感受野，未提“卷积”二字）

（3）降采样（Subsample）的平均池化层（Average Pooling）

（4）双曲正切（Tanh）的激活函数

（5）MLP作为最后的分类器

（6）使用卷积令层与层之间稀疏连接，减少了计算复杂性

不过，一般认为CNN的开山之作是Lecun 1989发表的另一篇论文《Backpropagation Applied to Handwritten Zip Code Recognition》不过这篇论文写的比较短，通篇没有找到CNN模型相关的原理解释说明。当时也没有提到“卷积”二字，而是被称为“感受野”。

不过从它的参考文献中，我们可以找到一篇重要的文章《1980-Fukushima-Neocognitron A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position》。这是日本人福岛邦彦，通过研究猫的视觉神经得到的启发，阐述了卷积和池化两个思想（当时还不叫卷积和池化）。

福岛邦彦的团队当时也开发了仿真模型Neocognitron（神经认知机），不过Neocognitron为什么没有发扬光大呢，感觉和CNN相比主要还是差在BP算法上，直到1998年Lecun提出了LeNet-5，把CNN推上了一个小高潮，之后Neocognitron也基本上放弃治疗了。

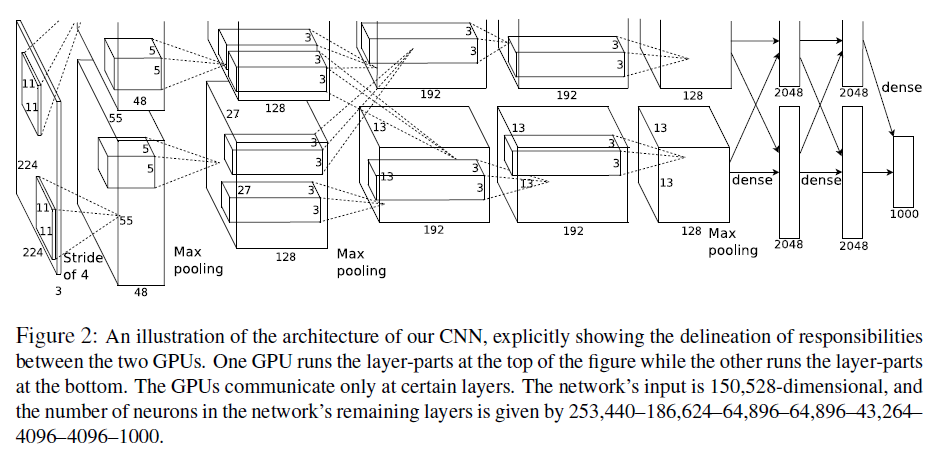
1998年到2012年世界在不断发生着巨大的变化，CNN也在默默的发展，其中有一篇论文要提一下，就是MIT脑科学研究中心的论文《Notes on Convolutional Neural Networks》，里面给了详细的CNN权值更新的公式。

2、AlexNet（2012）--王者归来

AlexNet在2012年ImageNet竞赛中以超过第二名10.9个百分点的绝对优势一举夺冠，这一突破是具有重大历史意义的，因为它使计算机视觉模型跨过了从学术demo到商业化产品的门槛，从而将深度学习和CNN的名声突破学术界，在产业界一鸣惊人。AlexNet的出现可谓是卷积神经网络的王者归来。

AlexNet的作者Alex Krizhevsky，是Geoffrey Hinton（杰弗里·欣顿）的博士生，Hinton也是人工智能三主神之一，在2018年获得了图灵奖。而Lucun曾是Hinton的博士后，所以说AlexNet和LeNet也可谓是一脉相承。

AlexNet为8层深度的CNN网络，其中包括5个卷积层和3个全连接层（不包括LRN层和池化层），如下图所示：



AlexNet的特点和贡献：

（1）使用ReLU作为激活函数，由于ReLU是非饱和函数，也就是说它的导数在大于0时，一直是1，因此解决了Sigmoid激活函数在网络比较深时的梯度消失问题，提高SGD（随机梯度下降）的收敛速度。

（2）使用Dropout方法避免模型过拟合，该方法通过让全连接层的神经元（该模型在前两个全连接层引入Dropout）以一定的概率失去活性（比如0.5），失活的神经元不再参与前向和反向传播，相当于约有一半的神经元不再起作用。在预测的时候，让所有神经元的输出乘Dropout值（比如0.5）。这一机制有效缓解了模型的过拟合。

（3）重叠的最大池化，之前的CNN中普遍使用平均池化，而AlexNet全部使用最大池化，避免平均池化的模糊化效果。并且，池化的步长小于核尺寸，这样使得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LRN（局部响应归一化）层，对局部神经元的活动创建竞争机制，使得响应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。

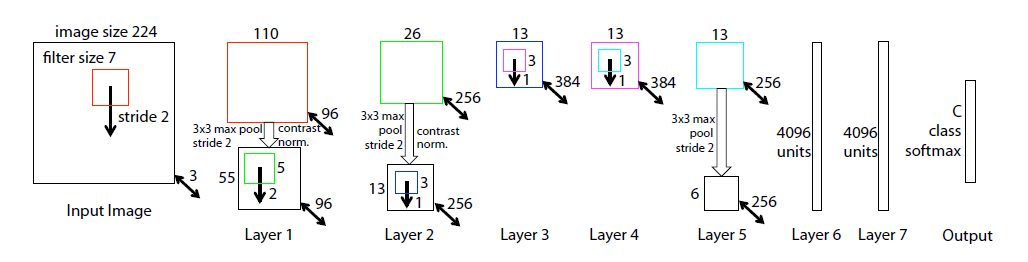
（5）使用GPU加速，加快了模型的训练速度，同时也意味着可以训练更大规模的神经网络模型。

（6）数据增强，随机从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），相当于增强了（256-224）\*（256-224）\*2=2048倍的数据量。使用了数据增强后，减轻过拟合，提升泛化能力。避免因为原始数据量的大小使得参数众多的CNN陷入过拟合中。

3、ZFNet（2013）--稳步前行

ZFNet是2013ImageNet分类任务的冠军，其网络结构没什么改进，只是调了调参，性能较Alex提升了不少。ZF-Net只是将AlexNet第一层卷积核由11变成7，步长由4变为2，第3，4，5卷积层转变为384，384，256。这一年的ImageNet还是比较平静的一届，其冠军ZF-Net的名堂也没其他届的经典网络架构响亮。

其论文是《Visualizing and Understanding Convolutional Networks》，网络结构如下：



4、VGG（2014 2rd）--越走越深

VGG-Nets是由牛津大学VGG（Visual Geometry Group）提出，是2014年ImageNet竞赛定位任务的第一名和分类任务的第二名的中的基础网络。VGG可以看成是加深版本的AlexNet. 都是conv layer + FC layer，在当时看来这是一个非常深的网络了，因为层数高达十多层，我们从其论文名字就知道了（《Very Deep Convolutional Networks for Large-Scale Visual Recognition》），当然以现在的目光看来VGG真的称不上是一个very deep的网络。

VGGNet探索了CNN的深度及其性能之间的关系，通过反复堆叠3\*3的小型卷积核和2\*2的最大池化层，VGGNet成功的构筑了16-19层深的CNN。

5、GoogLeNet（2014） Inception V1 V2 V3 V4，Inception-ResNet（2016）--大浪推手

Google Inception Net首次出现在ILSVRC2014的比赛中(和VGGNet同年)，以较大的优势获得冠军。那一届的GoogleNet通常被称为Inception V1，Inception V1的特点是控制了计算量的参数量的同时，获得了非常好的性能-top5错误率6.67%, 这主要归功于GoogleNet中引入一个新的网络结构Inception模块（该设计灵感来源于《Network in Network》，也就是NIN），所以GoogleNet又被称为Inception V1(后面还有改进版V2、V3、V4)架构中有22层深，V1比VGGNet和AlexNet都深，但是它只有500万的参数量，计算量也只有15亿次浮点运算，在参数量和计算量下降的同时保证了准确率，可以说是非常优秀并且实用的模型。

GoogLeNet的论文为《Going deeper with convolutions》，在这篇论文里，我看到了几个华人的名字，这是之前所没有的，说明大约从这个时候开始，华人科学家开始关注到CNN的发展。其中有一个名字有点儿熟悉，就是前段时间上了新闻的贾扬清（2019年3月份加入阿里巴巴硅谷研究院），他开发了著名的机器学习框架Caffe，并在Google参与设计和开发了Tensorflow。

特点：

（1）引入Inception结构，这是一种网中网（Network In Network）的结构，即原来的结点也是一个网络。

（2）中间层的辅助LOSS单元，GoogLeNet网络结构中有3个LOSS单元，这样的网络设计是为了帮助网络的收敛。在中间层加入辅助计算的LOSS单元，目的是计算损失时让低层的特征也有很好的区分能力，从而让网络更好地被训练。在论文中，这两个辅助LOSS单元的计算被乘以0.3，然后和最后的LOSS相加作为最终的损失函数来训练网络。

（3）后面的全连接层全部替换为简单的全局平均pooling，将后面的全连接层全部替换为简单的全局平均pooling，在最后参数会变的更少。而在AlexNet中最后3层的全连接层参数差不多占总参数的90%，使用大网络在宽度和深度上允许GoogleNet移除全连接层，但并不会影响到结果的精度，在ImageNet中实现93.3%的精度，而且要比VGG还要快。

6、ResNet（2015）--里程碑式创新

2015年何恺明推出的ResNet在ISLVRC和COCO上横扫所有选手，获得冠军。ResNet在网络结构上做了大创新，而不再是简单的堆积层数，ResNet在卷积神经网络的新思路，绝对是深度学习发展历程上里程碑式的事件。微软亚洲研究院将VGGNet发展为MSRANet，后来由何恺明及其同事进一步开发了ResNet

引入残差单元，解决退化问题。

7、Trimps-Soushen（公安三所），CUImage（商汤和港中文）（2016）--百花齐放

8、SENet、DenseNet（2017）--另寻蹊径

# 三、人工智能大神八卦

人工智能三主神（获得2018年图灵奖）：

Geoffrey Hinton 杰弗里·欣顿，证明BP算法

Yann Lecun扬·莱坎 是Geoffrey Hinton的博士后，LeNet作者，MNIST

Yoshua Bengio 约舒亚·本希奥，LeCun在AT&T贝尔实验室的同事，RNN，CIFAR

Alex Krizhevsky是Hinton的学生，发明AlexNet，在2012年ILSVRC（ImageNet Large Scale Visual Recognition Challenge）大赛上大放异彩，准确率比第二名高出10.9%以上，跨过了从demo到商用的门槛

华人大神：

吴恩达，在线教育平台Coursera，斯坦福精品课程，加盟百度

李飞飞，ImageNet数据库，加盟谷歌

何恺明 是 汤晓鸥在香港中文大学的博士生，ResNet作者

汤晓鸥是商汤科技的创始人和董事长

印奇，旷视科技的创始人，汤晓鸥的博士生

# 四、相关资料

1、【卷积神经网络发展历程】从LeNet、AlexNet到ResNet、SENet： https://blog.csdn.net/u012679707/article/details/80870625

2、CNN网络架构演进：从LeNet到DenseNet： https://www.cnblogs.com/skyfsm/p/8451834.html

3、ILSVRC竞赛详细介绍（ImageNet Large Scale Visual Recognition Challenge）： https://blog.csdn.net/sophia\_11/article/details/84570177

4、CNN（卷积神经网络）最早是哪一年提出，是如何发展的？： https://www.zhihu.com/question/47705441?sort=created

5、